Towards Isotope Effect Calculations in the Supramolecular Age

Drawing Conclusions from Model systems to Develop our Understanding of Large-Scale Interactions

Wilson, Philippe

Award date:
2017

Awarding institution:
University of Bath

Towards Isotope Effect Calculations in the Supramolecular Age:

Drawing Conclusions from Model systems to Develop our Understanding of Large-Scale Interactions

A Thesis submitted for the award of Doctor of Philosophy

Philippe B. Wilson

Department of Chemistry

University of Bath

November 2016

Supervisor: Professor Ian H. Williams

_____

Acknowledgements

I would first like to thank Ian, simply for everything. His guidance, forethought, support and understanding has been purely exceptional, and I am proud to have been his Ph.D student.

I'd also like to thank my better half Emma, for putting up with the grumpy days, when thesis writing was slow, and supporting me throughout it all, and of course, Kiddo. My parents, who have been there for me throughout everything, and continue to be rocks in all parts of my life, and Broggy, who was perfect for 14 years.

Table of Contents

Abstract

Isotope effects have been widely employed in the understanding of reactions, mechanisms and chemical change for decades. Herein is presented an evaluation of isotope effect theory in the context of contemporary computational studies. Model systems are used to illustrate the importance of considering the reacting system as part of the full environment, as opposed to a free subset of molecules, independent of the surroundings. B3LYP/aug-cc-PVDZ studies suggest the methyl cation exhibits unusual properties when exposed to a dielectric continuum, signifying that the UA0 continuum solvation cavity model can lead to erroneous results. Further B3LYP, M06 and MP2 studies on the anharmonicity of the C-H bond, and its influence on calculated vibrational frequencies and isotope effects show significant variation from benchmark values for the M06/6-31G+(d),aug-cc-PVDZ and aug-cc-PVTZ electronic structure methods. In combination with the polarised continuum model, anharmonic corrections applied to B3LYP/6-31+G(d) produces erroneous results, leading to the recommendation of avoiding these groupings of options within the same calculation method. A B3LYP/aug-cc-PVDZ study of an explicitly solvated 'cluster' system consisting of a methyl cation surrounded by five molecules of water, highlight the importance of interactions other than the commonly-considered donor-acceptor distances in methyl transfer reactions, in determining isotope effects and thereby, reactivity. Moreover, detailed kinetic isotope effect computations employing the recently redeveloped SULISO suite of isotope effect and vibrational characterisation software identify hydrogen bonding interactions perpendicular to the methyl transfer axis, as conceivably contributing to 3% variations recently used to support the compression hypothesis, and observed in isotope effects from experiment. The cutoff model, as postulated by Wolfsberg and Stern, and recently redeveloped by Williams, is applied to the methyl cation cluster system to identify degrees of freedom with significant contributions to the isotope effect, and therefore reactivity of the moiety. External degrees of freedom are seen to govern important contributions to the isotope effect, indicating that extreme care must be taken when applying the cutoff procedure to molecular Hessians. *Reductio ad absurdum* studies of cutoff levels show a major influence on the calculated isotope effects; the accuracy of the calculation of the reaction coordinate frequency and associated components significantly affect the accuracy of the

isotope effects obtained. The theory implicit in the SULISO suite of software is applied to alternative systems. Isotope effects for a ruthenium-catalysed C-H activation catalytic cycle are computed and used to deduce the nature of catalysis and the most influential structures within. C-H activation is found to proceed by formation of an agostic intermediate, before conversion and final dissociation of products, with the observed KIE of 2.22 agreeing well with the calculated value of 2.17. The impact of dispersion corrections (D3 Grimme dispersion with Becke-Johnson damping), and their alternative implementations is described, suggesting that care must be taken when employing dispersion in terms of energetic corrections as opposed to including structural considerations. Dispersion is then considered as an effect on isotope effects calculated from QM/MM Hessians on catechol-*o*-methyltransferase. Isotope effects themselves vary insignificantly between methods, however the structures of the QM region and energetics indicate a contribution from dispersion corrections when included in the optimisation routine. A set of guidelines for future calculations of isotope effects in large, supramolecular systems acts as a conclusion, unifying the outcomes of each study included herein.

Foreword

Isotope effects are some of the most subtle yet important probes into reaction mechanism and geometry change [1], yet some factors that affect their direction and magnitude were still largely unknown. A quick literature search for *kinetic isotope effect* can produce thousands of results, but how reliable is the data, how reliable are the calculations, and how reliable are the interpretations?

In 2014, it was postulated that it was necessary to explore some of the most tenacious questions in the field, by concentrating on a methodological investigation of the nature of isotope effects, and how they are affected by various aspects of the system they represent, such as environment and system size, whilst carrying out method validation on the electronic structure methods employed. After all, understanding interactions on a supramolecular scale begins first by understanding atomic and molecular factors.

Therefore, a series of projects designed to explore the basis of the current system and offer robust computational recommendations for the future calculation and interpretation of isotope effects was embarked upon, based on mainly small models as applied to supramolecular systems. Throughout the past few years, the effects of the environment,[2, 3] the vibrational description of the systems,[4] the level of inclusion of molecules within the calculation,[5] and the impact of the original electronic structure methods used to treat these systems have all been considered.

Additionally, the SULISO suite of programs has been applied to the calculation of not only organic and enzymic entities, but also organometallic reactions and catalytic cycles [6], showing the inherent transferability of the SULISO protocols within the chemical field.

A series of developmental concepts have been discovered, which have been presented to the community at conferences both in the UK and abroad. The effect of solvation on isotope effects formed the initial study. Previous discussion on the subject had resulted in some conjecture on the origins of solvent effects on isotope effects.

Through calculations of equilibrium isotope effects on a small system, with increasingly polar solvents, it was possible to confirm that solvation has a direct effect on the magnitude of isotope effects themselves.

Additionally, this work contributed to the investigation into United Atom Topological Model (UA0) insufficiencies when compared to the more recent Universal Force Field (UFF) model, all of which is described in Chapter 3.

Having established that the environment effectively contributed to isotope effect magnitude, this idea was developed into a more theoretical investigation into how the base data for isotope effect calculation; namely the Hessian and vibrational frequencies, are calculated. Looking more closely at anharmonic corrections and considering the benefits and detriments of these with respect to their harmonic equivalents, and experimental-harmonic scaling factors, it was concluded that the computational expense of anharmonic corrections certainly within the system studied, did not outweigh the advantages presented by harmonic scaling factors.

Furthermore, through the method validation stage of the investigation, a number of conflicts between electronic structure methods and solvation and/or vibrational calculation method were discovered. A full discussion of the work carried out is presented in Chapter 4.

Recognising that now certain ESMs and combinations of theoretical methodology yielded erroneous results, a reasonable method could be chosen to further investigate the effect of solvent, only this time explicitly. By creating a test system consisting of the same small molecule as the original environmental work, surrounded by a *cage* of water molecules at its axes; judicious use of internal coordinates allowed for the subtle manipulation of each solvent molecule, and the subsequent calculation of its effect on the isotope effect for the system.

Considering that the test molecule was chosen as a small-scale model for COMT methyl transfer, it was possible to qualitatively make comments on the results obtained, and their contribution to current hypotheses and conjecture in the field. The findings were of particular importance, and are fully discussed in Chapter 7.

As the effect of solvation both implicit and explicit was quite clear, the influence of molecular subsets on result reliability was important to consider. Using the same cage as a model system, numerous levels of cutoff were applied to the structure, and isotope effects calculated in order to gauge minimal and proximal cutoff regions. Introduced in Chapter 1 but developed in Chapter 6; the theory of cut-off Hessians and molecular

subsets is particularly important for large systems where QM/MM calculations and equivalents are applied. The investigation in Chapter 7 therefore particularly applies to the enzymic systems considered and provides important suggestions for research employing cutoff procedures for calculation efficiency.

With increases in computing power come the calculation of additional forces and interactions becoming computationally viable. Therefore dispersion as an effect was not often considered in early simulations as is often conceptually treated as only meaningful for particularly large systems.[7] A recent publication by Schreiner and co-workers revealed the surprising impact dispersion effects can have on numerous systems and system sizes.[8] Having rarely included these effects in our calculations previously, it was therefore timely to consider them explicitly for the enzymic systems in particular, and inorganic catalysis; the entire investigation being presented in Chapters 7 and 8.

Having surveyed a series of questions from the field, and providing answers to these, the final chapters conclude with summaries of the work, and guidelines for the calculation and interpretation of isotope effects for future studies, as presented in the context of this thesis. The outlook for isotope effect calculations, and the avenues that are currently being explored by other groups, internationally form the concluding chapter.

List of Abbreviations

A

AdoMet –*S*-adenosyl methionine

ADP – Adenoside diphosphate

AM1 – Austin Model 1, a semiempirical quantum chemical method

AMLA – Ambiphilic metal ligand activation

AMP – Adenosine monophosphate

a.m.u. – Atomic mass unit(s)

A_N – Association (mechanism)

B

B3LYP – Three-parameter hybrid functional based on Becke exchange with Lee, Yang, and Parr correlation

B_{CO} – Bond order (for C-O bond)

BIE – Binding isotope effect

BP86 – GGA functional based on Becke 1988 exchange and Perdew 86 correlation functionals

C

CCSD – Coupled cluster electronic structure theory with single and double excitations

CHARMM – Chemistry at Harvard Macromolecular Mechanics (molecular simulation program)

CMD – Concerted metallation deprotonation

COMT – Catechol-*O*-methyltransferase

CoA – Coenzyme A

D

DC – Dispersion corrected

DFT – Density functional theory

D_N – Dissociation (mechanism)

DO – Dispersion optimised

E

EA-VTST – Ensemble-averaged variational transition state theory

EIE – Equilibrium isotope effect

ESM – Electronic structure method

EX – contribution of population of excited vibrational states to an isotope effect from a molecular subset

EXC – contribution of population of excited vibrational states to an isotope effect

F

FAD – Flavin adenine dinucleotide

G

GD3BJ – Grimme D3 dispersion correction with Becke-Johnson damping

GGA – Generalised gradient approximation

H

HF – Hartree-Fock electronic structure theory

I

INT - Intermediate

IPFR – Isotopic partition function ratio

IRC – Intrinsic reaction coordinate

K

k_{cat} – rate of conversion of substrate by enzyme (turnover number)

KIE – Kinetic isotope effect

K_m – Michaelis constant

L

LACVP – basis set consisting of 6-31G (H-Ar) and LANL2DZ (heavy atoms) core potentials

LDA – Local density approximation

LDH – Liver alcohol dehydrogenase

M

M06 – 2006 Minnesota density functional from the Truhlar group

MADH – Methylamine dehydrogenase

MC – Monte Carlo

MD – Molecular dynamics

Me - Methyl

MNDO – Modified neglect of diatomic overlap (semiempirical electronic structure method)

MI – Moment of inertia term contributing to the isotope effect calculated from a molecular subset

MM – Molecular mechanics

MMI – Contribution of the mass and moment of inertia term to a calculated isotope effect

MP2 – Second order Møller-Plesset perturbation theory

N

NAD+/NADH – Nicotinamide adenine dinucleotide

O

ONIOM – Our own n-layered integrated molecular orbital and molecular mechanics

OPLA – Out of plane large amplitude molecular vibration

P

PCM- Polarised continuum model

p-cym – *para*-cymene

PES – Potential energy surface

PF – Partition function

PMF – Potential of mean force

PS – Product structure

Q

QM – Quantum mechanics/mechanical

QM/MM – Quantum mechanics/ molecular mechanics

R

r_{ax} – axial C-O distance in cage structures

RECP – Relativistic effective core potential basis set

r_{eq} – equatorial C-H···O distance in cage structures

RMS – Root mean square

RS – Reactant structure

S

SAM – S-adenosylmethionine

SCC-DFTB – self-consistent charge density functional tight binding quantum chemical method

scis – Scissor vibration between three atoms in a molecule

SCT – Small curvature tunnelling

SDD – Dunning/Huzinga full double zeta basis set (H - Ar) with Stuttgart/Dresden ECPs on remainder

SMD – Solvation model density (continuum solvation model)

S_N1 –Nucleophilic substitution, unimolecular in rate determining step

S_N2 –Nucleophilic substitution, bimolecular in rate determining step

SULISO – Sulis (Bath) Isotope suite of calculation software

T

THF – Tetrahydrofuran

TIP3P – Three site transferrable intermolecular potential

TS – Transition structure

T_x – Translation in the *x* direction

T_y – Translation in the *y* direction

T_z – Translation in the *z* direction

U

UA0 – United atom topological cavity model

UFF – Universal force field cavity model

umbr – Umbrella vibration

V

VCI – Vibrational configuration interaction theory

V_{max} – Maximum velocity of an enzyme reaction; the rate of reaction at complete active site saturation

VP – Vibrational product

VPT – Vibrational perturbation method

Z

ZP – Zero point energy contribution to the isotope effect calculated from a molecular subset

ZPE – Zero point energy

List of Publications

6. **P.B. Wilson** and I. H. Williams, *Software:X,* **2016**, *6*, pp 1-6

The SULISO Suite of Vibrational Characterisation and Isotope Effect Calculation Software

DOI:10.1016/j.softx.2016.11.001

In Thesis: *Chapter 2*

5. **P. B. Wilson** and I. H. Williams, in *Simulating Enzyme Reactivity*, ed. I. Tunon and V. Moliner, **2016**, *Royal Society of Chemistry Publishing*, pp 150-184
Chapter 5: Kinetic Isotope Effects

In Thesis: *Chapters 1, 6, and 7.*

4. J. Leitch, **P. B. Wilson**, *et al.*, *ACS Catalysis* , **2016***, 6, pp 5520-5529*
Ruthenium(II) Catalyzed C-H Functionalization Using the Oxazolidinone Heterocycle as a Weakly Coordinating Directing Group: Experimental and Computational Insights
DOI: 10.1021/acscatal.6b01370

In Thesis: *Chapter 8*

3. **P. B. Wilson** and I. H. Williams, *Angew. Chem. Int. Ed.*, **2016**, *55*, pp 3192-3195**/** *Angew. Chem.*, **2016**, *128*, pp 3244-3247
Equatorial CH···O Interactions Influence Secondary Kinetic Isotope Effects for Methyl Transfer
DOI: 10.1002/anie.201511708 **/** 10.1002/ange.201511708
In Thesis: *Chapter 5*

2. **P. B. Wilson**, and I. H. Williams, *Mol. Phys.*, **2015**, *113* (13-14), pp 1704-1711
Critical evaluation of anharmonic corrections to the equilibrium isotope effect for methyl cation transfer from vacuum to dielectric continuum
DOI: 10.1080/00268976.2015.1007106

In Thesis: *Chapter 4*

1. **P. B. Wilson**, P. J. Weaver, I. R. Grieg and I. H. Williams, *J. Phys. Chem. B.*, **2015**, *119* (3), pp 802-809
Solvent Effects on Isotope Effects: Methyl Cation as a Model System
DOI: 10.1021/jp505344a

In Thesis: *Chapter 3*

1. Review of Isotope Effects for enzyme-catalysed reactions

This chapter consists of an introduction to isotope effect studies, based on work published in *Simulating Enzyme Reactivity*, by Wilson & Williams, edited by Moliner and Tunon. Methods for calculation of kinetic isotope effects and their application to reactions catalysed by enzymes are included, with consideration of cut-off approximations, empirical bond-order – bond-energy approaches, quantum-mechanical cluster methods, and hybrid quantum-mechanics/molecular mechanical methods. The commonality between these is the use of a Hessian matrix of force constants within the harmonic approximation. The merits of methods appropriate for supramolecular systems, including enzymes, are discussed in contrast to commonly used methods developed for gas-phase molecules, and the importance of averaging in conformationally flexible systems is emphasised.

1.1 Introduction

Enzymes are proteins, which each act as catalysts for specific types of chemical/biochemical reaction. The reactants for these processes are described as substrates, as interact in a complex fashion with the enzyme and its reaction centre, the active site. Enzymes have by nature, a high specificity, as can catalyse only reactions which their active site geometry or electrostatic topology permit. Some enzyme-catalysed reactions are examples of zero-order reactions, or reactions where the rate or speed of the reaction, depends on the concentration of reactant, or *substrate* in this case. Active sites on enzymes can bind to only one unit of substrate at any one time, regardless of the concentration of substrate present. A simple model to describe enzyme catalysis refers to its similarity with a lock and key. In this model, the substrate molecules fit in the active site of the enzyme similarly to a key fitting in a lock. As such, when the reaction has completed and products are formed, these no longer fit in the active site area and they therefore dissociate with the enzyme. The enzyme can then catalyse another reaction. In this model, it is noted that the enzyme is not permanently changed by this procedure, and in simple terms, is recyclable.

Enzyme action tends to be described by Michaelis-Menten kinetics. Indeed, within this formulation of the rate law, the rate of the enzyme-catalysed reaction depends on the concentration of the enzyme, ***E***, itself. Here, substrates, ***S***, are converted into products, ***P***:

$$E + S \rightleftharpoons ES \rightarrow P + E \tag{1}$$

Where ES is the enzyme-substrate complex.

Assuming that:

$$[ES] = \frac{k_a[E][S]}{k_a' + k_b} \tag{2}$$

Where k_a is the rate constant for the forward reaction step to form the ES complex, k_a' represents the rate constant for the reverse step of ES to E + S, and k_b the rate constant for formation of the products and unbound enzyme. This is based on the prototypical pre-equilibrium step where the ES complex is formed:

$$K = \frac{[ES]}{[E][S]} = \frac{k_a}{k_a'} \tag{3}$$

The rate of formation of the product is therefore dependent on the formation of the ES complex:

$$\frac{dP}{dt} = k_b[ES] \tag{4}$$

The formation of the ES complex thus describes the arrival at the equilibrium step. In order to obtain an expression for the ES and describe the rate in terms of [E] and [S], the equilibrium state of the ES complex can be shown as:

$$\frac{d[ES]}{dt} = k_a[E][S] - k'_a[ES] - k_b[ES] \approx 0$$

(5)

This yields the net rate of change of ES with time. Equation 5 is solved to provide an expression for the concentration of ES:

$$[ES] \approx \frac{k_a[E][S]}{k'_a + k_b}$$

(6)

The rate of formation of products (Equation 6) can now be rewritten as:

$$\frac{dP}{dt} \approx k[E][S]$$

(7)

where

$$k = \frac{k_a k_b}{k'_a + k_b}$$

(8)

As [E] and [S] are the concentrations of the free enzyme and substrate, the total concentration of enzyme, $[E]_0$ is simply [E] + [ES]. As there is only a small amount of enzyme added to the reaction, the concentration of free substrates remains at approximately the same level as the total substrate concentration, and [S] can therefore be assumed to be approximately equal to $[S]_{total}$.

$$[ES] = \frac{k_a[E]_0[S]}{k'_a + k_b + k_a[S]}$$

(9)

The rate of formation of product, ***P***, is therefore:

$$\frac{dP}{dt} = k[E]_0 \tag{10}$$

where:

$$k = \frac{k_b[S]}{K_M + [S]} \tag{11}$$

and:

$$K_M = \frac{k'_a + k_b}{k_a} \tag{12}$$

where K_M is the Michaelis Constant.

Equation 10 suggests that the rate of enzyme catalysis is proportional to the enzyme concentration, and more intricately related to the substrate concentration. However, when the concentration of substrate, *[S]*, is far greater than the Michaelis constant, equation 10 can be expressed simply as equation 13:

$$\frac{dP}{dt} = k_b[E]_0 \tag{13}$$

The rate of the reaction catalysed by the enzyme is then zero order for the substrate. This indicated that under the appropriate conditions, [S] is high so that regardless of products being formed, it will remain at approximately the same level throughout. This suggests that the maximum rate of product formation (V_{max}) is achieved when the enzyme is saturated with substrate:

$$V_{max} = k_b[E]_0 \tag{14}$$

The rate of product formation, k_b, is referred to as the turnover number, or rate at which products are formed. This therefore successfully describes the manner in which enzyme catalysis occurs. Enzymes bind to the substrates and catalyse the reactions within the active site before forming products. Most usually, enzymes contribute to lowering the activation energy, or stabilizing the transition structure of a chemical reaction. This is an important part of enzyme catalysis and accurate descriptions of transition structures produce reliable kinetic isotope effects for chemical change.

Indeed, inherent in the definition of isotope effects, particularly Kinetic, are the ideas of transition state theory (TST). TST is based on the idea that chemical reactions involve the breaking and formation of bonds, which engender changes in potential energy. Let us consider the following single step reaction under standard conditions:

$$A + B_2 \rightarrow AB + B$$

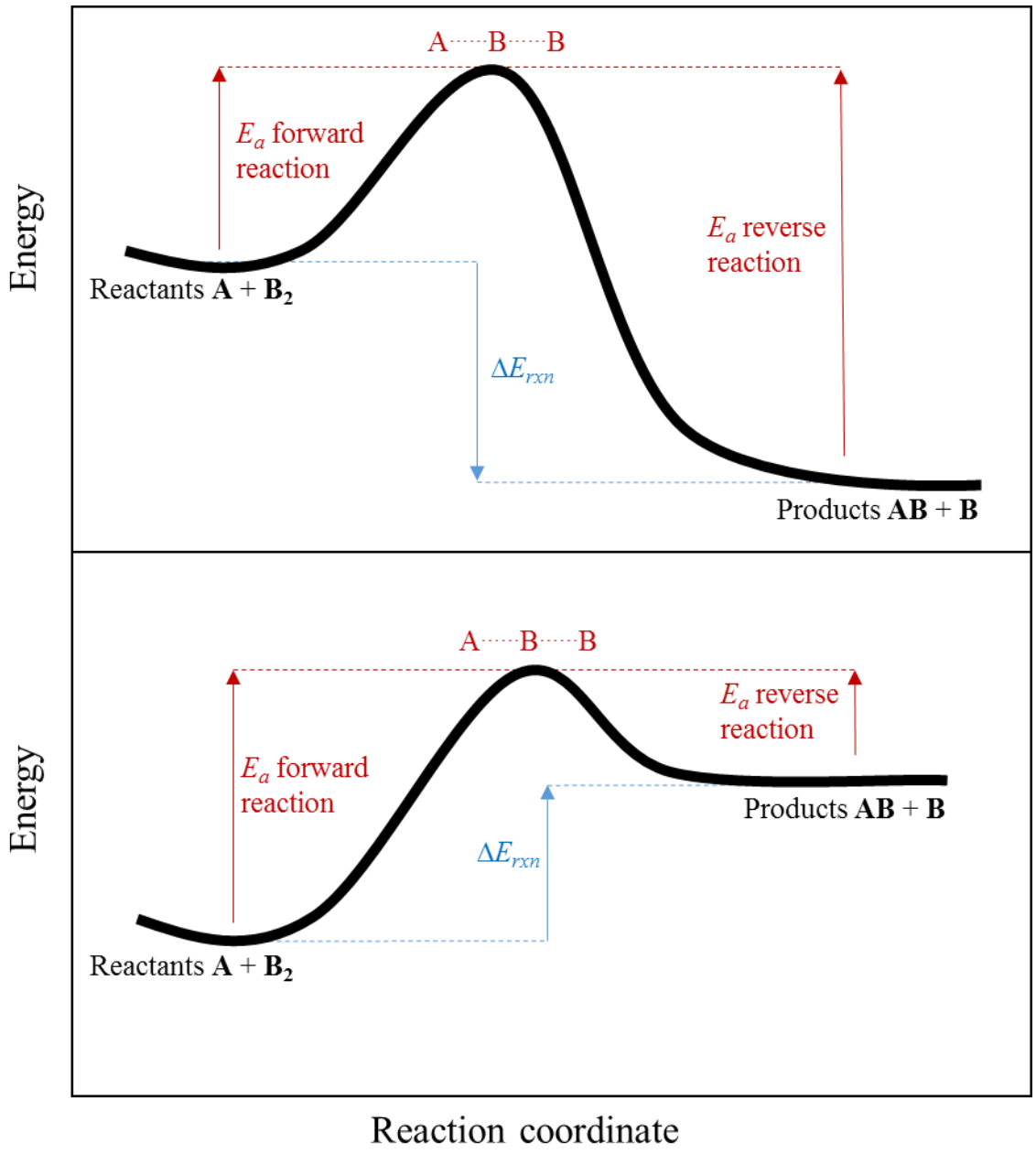


Figure 1. Potential energy diagram for exothermic (top diagram), and endothermic (bottom diagram) reactions. The reaction coordinate represents the progress of the reaction. E_a represents the activation energy, or the kinetic energy required by reactant molecules to achieve the higher energy state. ΔE_{rxn} is the net change in energy on going from reactants to products.

In the first part of Figure 1, the ground state energy of reactants A + B_2 is higher than that of the product ground state. The difference between these two energies represents the energy release on product formation. Conversely, in the second part of Figure 1, the ground state energy of the reactants is lower than that of the products, as such energy is not released but absorbed in order to form the products. In most chemical reactions, covalent bonds tend to be broken in order for other to form. This tends to occur through intermolecular collisions, where the molecules need to possess sufficient kinetic energy to overcome the stabilization due to potential energy of the covalent bonds. The principle of transition state theory (TST) states that reactants must pass through a short-lived, high energy intermediate state, or transition state (TS), before forming products. This TS is denoted *A·····B·····B* in Fig. 1.

In order for the reverse reaction to occur, product molecules must have sufficient energy to overcome the activation barrier (E_a *reverse*) in order to reach the TS. External factors have an influence on the nature of reactions. Indeed, increasing the temperature changes the rare of formation of products by changing the number of molecules with enough energy to overcome the barrier to TS formation. Likewise, introduction of a catalyst often provides an alternative pathway for a reaction, with a lower activation energy.

The reaction illustrated in Fig. 1 shows a single-step process. However in reality, most reactions are formed of a number of successive steps together forming a mechanism. In these stepwise mechanisms, one step has a tendency to be slower than others, therefore the overall reaction is limited to be only as fast as its slowest elementary reaction step. Often called the rate-determining step, this part of the reaction mechanism becomes the defining part for the speed of the entire reaction mechanism. In order to form a balanced equation for the whole reaction, there must be a sum of the individual elementary steps, therefore for the overall reaction:

$$w\mathrm{A} + x\mathrm{B} \rightarrow y\mathrm{C} + z\mathrm{D}$$

where *w , x, y,* and *z* represent the stoichiometry of the reactants and products.

The rate law derived from experiment can be expressed as:

$$\text{rate} = k[\text{A}]^n[\text{B}]^m$$

where n and m are related to the coefficients of the reactants in the rate determining step. From experimental data, it is often possible to postulate a mechanism by which reactions can occur. These mechanisms must be consistent with experimental data, and a number of parameters derived from experiment, such as kinetic and equilibrium isotope effects, can be used to describe the possible mechanism of a reaction.

A kinetic isotope effect (KIE) is a ratio of rate constants deviating from unity for a pair of reactions under identical conditions except for the isotopic composition of the reacting species, and is a consequence of the quantum-mechanical nature of nuclear motion. Historically, the theory of KIEs for reactions of gas-phase molecules was developed within the framework of transition state theory by the methods of statistical mechanics in terms of partition functions evaluated within the Born-Oppenheimer, harmonic-oscillator, rigid-rotor and ideal-gas approximations.[9, 10] A partition function can be defined as the ration of total number of molecules, to those in the ground state of a system. In the context of the studies within this thesis, an isotopic partition function ratio (IPFR) is a ratio of molecular partition functions ('heavy/light') for a pair of isotopologous species.

Neglecting tunnelling and transmission-factor contributions, the ratio of rate constants k/k' (where the prime denotes the heavier isotopologue) may be expressed as the product of three factors (Equation 15) involving translational, rotational and vibrational partition functions for the isotopologous reactant-state (RS) and transition-state (TS) species.[11, 12] The first is the mass and moments-of-inertia factor (MMI), where $|\boldsymbol{I}|$ is the determinant of the moment-of-inertia tensor (i.e. the product of the three principal moments of inertia) of a non-linear N-atomic molecule of molecular mass M; rotational symmetry factors are omitted here since they are generally equal to unity for applications to enzymatic reactions. The second is the Boltzmann excitational (EXC) factor corresponding to the relative populations of the higher vibrational quantum states, and the third is the zero-point energy (ZPE). EXC involves a product, and ZPE a summation, over $3N - 6$ and $3N - 7$ real vibrational frequencies of RS and TS, respectively, where $u = hc\nu/k_BT$ with frequency (s^{-1}) replaced by wavenumber ν (cm^{-1}) and

h, k_B, c and T are the Planck and Boltzmann constants, the velocity of light and the absolute temperature.

$$k/k' = \frac{\left[\frac{|I'|}{|I|}\right]_{RS}^{1/2}\left[\frac{M'}{M}\right]_{RS}^{3/2}\prod_i^{3N-6}\left[\frac{[1-\exp(-u_i)]}{[1-\exp(-u_i')]}\frac{\exp(u_i/2)}{\exp(u_i'/2)}\right]_{RS}}{\left[\frac{|I'|}{|I|}\right]_{TS}^{1/2}\left[\frac{M'}{M}\right]_{TS}^{3/2}\prod_i^{3N-7}\left[\frac{[1-\exp(-u_i)]}{[1-\exp(-u_i')]}\frac{\exp(u_i/2)}{\exp(u_i'/2)}\right]_{TS}} \qquad (15)$$

$$\mathrm{KIE} = \underbrace{\mathrm{MMI}}\ \times\ \underbrace{\mathrm{EXC}}\ \times\ \underbrace{\mathrm{ZPE}}$$

The MMI, EXC and ZPE factors in Equation 15 are derived from the partition functions for the molecular species involved. Cancellation of physical constants, volume and temperature common to the translational and rotational partition functions for all four species leads to the MMI factor as a ratio of 'reduced' partition functions for isotopologous species.[9, 13] EXC describes the varying population statistics of harmonic vibrational energy levels, by means of vibrational partition functions evaluated with respect to the lowest vibrational energy level, and ZPE is, of course, obtained simply from the differences between that quantised energy level and the potential energy at a classical stationary point. It is common to replace MMI in Equation 15 by an equivalent factor containing only vibrational frequencies. According to the Teller-Redlich product rule,[14] the masses and moments of inertia for a pair of isotopologues are related to the vibrational frequencies: the equality expressed by Equation 16 assumes separability of translational and rotational motions from vibrational motions within the harmonic approximation. Substituting the vibrational product VP for each of RS and for TS in Equation 15, the KIE may be written as Equation 17, the Bigeleisen equation, where VPR is the vibrational product ratio.[9, 15] Two points should be noted: (i) the product of ratios of atomic masses m that appears in Equation 16 vanishes from the KIE because it is identical for both RS and TS; (ii) it is conventional to consider the ratio of imaginary transition frequencies for the TS as a separate factor, so that (in Equations 17 and 18) the VPR and EXC products and the ZPE summation are all taken over 3N – 7 real TS frequencies.

$$\prod_{i}^{3N-6} \frac{\nu_i'}{\nu_i} = \mathrm{VP} = \mathrm{MI} = \left[\frac{|\boldsymbol{I}'|}{|\boldsymbol{I}|}\right]^{1/2}\left[\frac{M'}{M}\prod_{j}^{N}\frac{m_j}{m_j'}\right]^{3/2} \tag{16}$$

$$k/k' = (\nu_\ddagger / \nu_\ddagger') \times \mathrm{VPR} \times \mathrm{EXC} \times \mathrm{ZPE} \tag{17}$$

$$\prod_{i}^{3N-6} \frac{\nu_i'}{\nu_i} = \mathrm{VP} = \mathrm{MI} = \left[\frac{|\boldsymbol{I}'|}{|\boldsymbol{I}|}\right]^{1/2}\left[\frac{M'}{M}\prod_{j}^{N}\frac{m_j}{m_j'}\right]^{3/2} \tag{18}$$

$$\mathrm{VPR} = \prod_{i}^{3N-6}\left[\frac{\nu_i'}{\nu_i}\right]_{\mathrm{RS}} \Big/ \prod_{i}^{3N-7}\left[\frac{\nu_i'}{\nu_i}\right]_{\mathrm{TS}}$$

Use of the Bigeleisen equation (written more compactly as Equation 19) enables the determination of KIEs from normal-mode frequencies without the need to consider the MMI factor explicitly; this fact has been exploited in several computer codes for KIE calculations.[16-18]

$$\mathrm{KIE} = \frac{\prod_{i}^{3N-6}\left[\dfrac{\nu_i \sinh(\frac{1}{2}u_i)}{\nu_i' \sinh(\frac{1}{2}u_i')}\right]_{\mathrm{RS}}}{\prod_{i}^{3N-7}\left[\dfrac{\nu_i \sinh(\frac{1}{2}u_i)}{\nu_i' \sinh(\frac{1}{2}u_i')}\right]_{\mathrm{TS}}\left[\dfrac{\nu_\ddagger'}{\nu_\ddagger}\right]_{\mathrm{TS}}} \tag{19}$$

The ratio of imaginary transition frequencies in Equations 17 and 19 is a temperature-independent factor which arises because motion in the direction of the transition vector is considered to be a classical translation in transition state theory. The other factors are purely quantal and their values tend to unity in the limit of very high temperature. The products over real vibrational frequencies in Equation 15, which determine the EXC and ZPE factors, correspond to ratios of quantum corrections to the classical vibrational partition functions. It is perfectly logical and consistent to include also the quantum correction to the classical partition function for motion in the TS vibrational mode with the imaginary frequency. Indeed, Bigeleisen did this, following Wigner's approximate

treatment,[19] in his original formulation of KIE theory.[9] Application of a quantum correction to every degree of freedom of both RS and TS is equivalent to consideration of both zero-point energy and tunnelling. Assuming that motion along the reaction coordinate in a TS is separable from the other vibrational degrees of freedom, then use of the simplified form of Bell's treatment for tunnelling through a one-dimensional parabolic barrier [20] gives a particularly elegant expression for the KIE (Equation 20) which involves $3N-6$ harmonic vibrational frequencies for both the RS and TS.

$$\mathrm{KIE} = \frac{\prod_{i}^{3N-6} \left[\frac{\nu_i \sinh(\tfrac{1}{2} u_i)}{\nu_i' \sinh(\tfrac{1}{2} u_i')}\right]_{RS}}{\prod_{i}^{3N-7} \left[\frac{\nu_i \sinh(\tfrac{1}{2} u_i)}{\nu_i' \sinh(\tfrac{1}{2} u_i')}\right]_{TS} \left[\frac{\nu_\ddagger' \sin(\tfrac{1}{2} u_i)}{\nu_\ddagger \sin(\tfrac{1}{2} u_i')}\right]_{TS}} \qquad (20)$$

These equations provide the starting point for discussion of computational simulation of KIEs for enzymatic reactions in the rest of this chapter. Where equilibrium isotope effects (EIEs) are the subject, the TS part of equations 15 to 20 are replaced by parameters for product structures, or PS. Indeed, one factor present in KIE as opposed to EIE computations is the tunnelling factor. KIEs are a subtle probe for the effect of tunnelling in reactions. Indeed, tunnelling occurs when a particle passes through an energy barrier as opposed to proceeding over it. As tunnelling involves crossing the barrier, as opposed to overcoming it, the process is particularly sensitive to both the width of the barrier, and the mass of the tunnelling particle. Indeed, the wider the barrier, and heavier the particle the more difficult the tunnelling process. Huge KIE values can be attributed to tunnelling; proton transfers are often subject to tunnelling effects, due to the small mass of the tunnelling particle. Moreover, reactions in excited vibrational states tend to have barrier of lower width than ground state reactions, thereby increasing the probability of particle tunnelling. Tunnelling reactions tend to be characterised by KIEs increasing with decreasing temperature, largely negative entropies of activation due to effectively ordered TSs, and activation energy differences for isotopologues larger than the difference in their zero point energies. Overall, the factors responsible for tunnelling can be observed in the non-linear Arrhennius

behaviour of rate constants. Generally, temperature ranges in experiments are not extended enough to observe these changes in Arrhenius plots.

The Bell correction is often implemented to account for tunnelling in calculations of KIEs. It assumes that the reaction barrier is affected by a group of particles with a Boltzmann distribution of energies; an assumption which has been suggested to be unrealistic. Indeed, within the Bell tunnelling model, a bound particle in a distinct state is not accounted for. Moreover, the Bell assumption of a "continuum of vibrational energies for reactant and product hydrogen bonds"[21] is likely to be lacking. For enzymic reactions, barriers tend to be relatively low, therefore C-H bonds would tend to be vibrating in the ground or first vibrational excited state. Nonetheless, the Bell correction has proven remarkably reliable for a number of studies. Klinman has noted that k_D/k_T and k_H/k_T deviations were successfully predicted through use of the Bell correction for the yeast alcohol dehydrogenase reaction, before experiment verified these findings.

1.2 The Cutoff Approximation

Use of the equations above requires separate Hessian matrices for the RS and for the TS in order to obtain harmonic frequencies of vibration, but the same Hessian may be employed with different atomic masses for each member of an isotopologous pair. However, to describe an entire system of enzyme and substrate(s), the large number of atoms necessarily implies very large Hessians for RS and TS. Early applications of the Bigeleisen equation, even to reactions of simple organic molecules, were restricted by the practical limitations of computer memory and processing power available in those times so that it was often necessary to employ truncated molecular models containing fewer atoms.

Furthermore, there existed another very serious problem that, in effect, also limited the size of system that could be treated in KIE computations. The advantage of using the Bigeleisen equation involving only vibrational frequencies is that, in principle, these could be obtained from experimental observations for isotopologous species. In practice,

for most molecules of interest, there were insufficient data to allow solution of the problem. Even if observed fundamentals were available, the Bigeleisen equation required harmonic frequencies which typically were not available. Moreover, no experimental frequencies were available at all for TSs!

Much effort in the 1960s (and later) went into vibrational analysis to extract force constants from spectra for stable molecules and, with the assumption of transferability, to construct Hessians for analogous species. The computer programs in use then[22] were implementations of the FG matrix method [23] employing a valence-coordinate representation of the molecule (in terms of bond stretching, angle bending, etc.) involving $3N-6$ non-redundant internal coordinates for a non-linear N-atomic molecule. The matrix of force constants was usually constructed to contain only diagonal elements, each corresponding to an individual bond or angle for which a value could be chosen by analogy with similar coordinates in other molecules, with the exception that a small number of off-diagonal elements were included for a TS with values carefully engineered in order to obtain a transition vector with an imaginary frequency.[24]

Wolfsberg and Stern [15] noted that it was possible to leave out parts of a 'large' molecule without significantly affecting the value of a calculated KIE. A "cut-off" that excluded all atoms more distant from a site of isotopic substitution than three bonds was found to yield isotope effects negligibly different from those obtained from a "full" model system, and they concluded that a 3-bond cut-off was always a justifiable procedure at room-temperature and above. Using models comprising only diagonal elements in the valence force-constant matrix, non-negligible KIEs were found only in cases with significant changes in force constants for valence coordinates directly involving an isotopically-substituted atom as between RS and TS. A 2-bond cut-off was usually justifiable, as this would include any force-constant changes in both stretching and angle-bending coordinates, despite excluding torsional coordinates involving the site of isotopic substitution as a terminal atom defining a dihedral angle. The other requirement for the cut-off procedure to work was that that values of force constants retained in the cut-off model must be the same as those for corresponding valence coordinates in the full model. Melander and Saunders discussed the practical usefulness of this simplification, as a means by which to reduce the computational cost of isotope-effect calculations, but added a further wise precautionary note that the cut-off procedure would be valid only

if it were justifiable to ignore the influence of medium effects on the isotope effect.[24] In addition to making computations more tractable by reducing the dimension of the force-constant matrix **F** in valence coordinates (or of a matrix of non-redundant symmetry force constants constructed from redundant valence force constants), the cut-off procedure reduced the number of coordinates for which force-constant values needed to be assigned. In view of the paucity of reliable data, this was a definite benefit.

1.3 The BEBO Vibrational Analysis Method for KIE Calculations

An alternative method for computational solution of the vibrational secular equations was proposed by Gwinn,[25] which simply involved diagonalization of the mass-weighted Hessian matrix **V** in Cartesian coordinates. The Hessian could still be constructed from valence force constants, as before, but any (immediate) problem with redundant valence coordinates was removed, and the $3N$ eigenvalues of **V** corresponded (in general) to $3N$ – 6 vibrations and 6 zero-valued translations and rotations. Although the dimension of the matrix to be diagonalised was larger (by 6) than in the **FG** method, increases in computational power soon made this an irrelevant point. Essentially all harmonic vibrational frequency calculations are now performed by this method.

Gwinn's method was incorporated into a computer program (QREL)[26] for calculations of relative rate constants including (but not limited to) KIEs written by Sims in 1975 during a sabbatical in Sheffield which was the precursor to the well-known BEBOVIB-IV program[27] for KIE calculations (as well as to the IHW's CAMVIB and CAMISO programs).[28, 29] BEBOVIB utilised Johnston's Bond-Energy–Bond-Order method,[30] together with empirical relationships between bond order, bond length and force constants, in order to provide estimates for otherwise unknown geometrical and force-constant parameters for TSs.[27, 31] The details of the method as applied to calculations of KIEs for enzymic reactions have been described very thoroughly by Berti.[32]

In what appears to have been the first example of KIE calculations for an enzymic reaction, Schowen and co-workers applied the BEBOVIB program to modelling of KIEs for the reaction catalysed by the enzyme catechol *O*-methyltransferase (COMT).[33] Their

study was initiated in 1978 and, in view of its seminal importance, it is worth reviewing its key features here. A 2-bond cut-off was employed to construct a 17-atom truncated structure (Figure 2) for which RS and TS geometries and valence force constants were derived from literature values for analogous coordinates together with simple relationships based upon the Pauling bond orders B_{CO} and B_{CS} of the making and breaking bonds between the carbon atom of the transferring methyl group and the oxygen and sulphur atoms of the nucleophile and nucleofuge, respectively. In principle the structural space containing the TSs was continuously variable. In practice, using these relations, a 2D grid of TSs was constructed and the primary (1°) ^{13}C, ^{18}O and ^{34}S and secondary (2°) $^{2}H_3$ KIE was computed as a function of the parameters B_{CO} and B_{CS} at each point on the grid, and a map of contours of constant KIE was plotted for each isotopic substitution (Figure 2). Experimental values for the ^{13}C and $^{2}H_3$ KIEs had been measured and could be compared with the computed maps: consideration of the values of the mean of the experimental KIE, plus and minus one standard deviation, enabled a range of TS structures to be delineated that were consistent with each KIE. Superimposition of the two maps generated smaller ranges of allowed TS structures that were consistent with both the ^{13}C KIE and the $^{2}H_3$ KIE observed for the COMT-catalysed reaction and separately for a model uncatalysed reaction in solution. These results indicated that the TS for the enzyme-catalysed reaction might have partial bond orders B_{CO} and B_{CS} significantly larger than those in the TS for the uncatalysed reaction, which might correspond to these bonds each being about 0.15 Å shorter in the enzymic TS, consistent with the idea that the enzyme's catalytic power might arise from compression of the S_N2 TS.

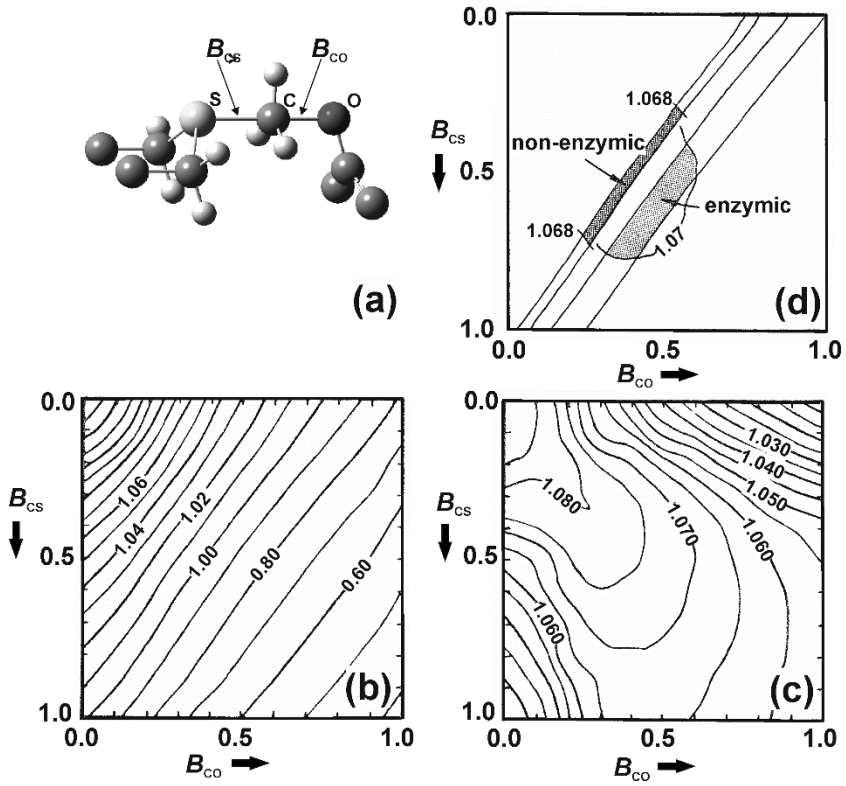


Figure 2 (a) Two-bond cut-off model for transition state of methyl transfer between sulfur of AdoMet and oxygen of catecholate used in the BEBOVIB modelling of Rodgers *et al.*[33] (b) Map of calculated 2° $^{2}H_3$ KIEs as a function of bond orders B_{CO} and B_{CS} to nucleophile and nucleofuge. (c) Similar map of calculated 1° ^{13}C KIEs. (d) Superposition of (b) and (c) showing regions consistent with experimental KIEs for non-enzymic (dark shading) and enzymic (light shading) methyl transfer. Graphic Adapted with permission from J. Rodgers, D. A. Femec and R. L. Schowen, *J. Am. Chem. Soc.* 1982, **104**, 3263–3268.[33] Copyright (1982) American Chemical Society.

Amongst enzymologists, the BEBOVIB approach was subsequently taken up by Hermes and Cleland to model the glucose-6-phosphate hydrogenase reaction,[34] and then most notably by Schramm and co-workers in a series of papers over many years, beginning with their study of adenosine monophosphate (AMP) nucleosidase[35, 36] and continuing with applications to nucleoside hydrolysis,[37] AMP-deaminase,[38] purine nucleoside phosphorylase,[38] NAD+ hydrolysis,[39-41] and ADP-ribosylation.[42, 43] Uses of BEBOVIB by

others include applications to glucoside hydrolysis,[44] orotate phosphoribosyltransferase,[45] oxalate decarboxylase[46] and adenosine nucleosidase.[47]

The isotopic mapping approach of using BEBOVIB in combination with multiple experimental KIE measurements for different isotopic substitutions, pioneered by Schowen and co-workers,[33] was formalised by Berti and Schramm[41] as the 'structure interpolation method'. Manual adjustment of bond-order parameters to give a TS structure yielding calculated KIEs that agree with experimental values is not only laborious but also error-prone and unable to predict possible alternative TS structures. A systematic procedure to use BEBOVIB calculations to explore many TS structures within the reaction space was first demonstrated by Sims and co-workers for a non-enzymic reaction,[48] but Berti and Schramm's protocol[49] used either X-ray crystal structures or QM optimised structures for RS and TS, and interpolations in all the internal coordinates are made without recourse to the empirical relationships used in the early work. An alternative for BEBO-style vibrational analysis and isotope effect calculations was Huskey's VIBIE program.[49]

Although when used carefully within a systematic procedure, the BEBOVIB method was capable of providing valuable insight concerning TS structure leading, for example, to successful design of TS-analogue enzyme inhibitors as potential drugs,[50] the fundamental lack of data for TS geometries and force constants was always a concern to some at least.

1.4 QM cluster calculations of KIEs

1.4.1 Early examples

As developments in semi-empirical molecular orbital theory began to allow for calculations of vibrational frequencies for TSs, and thereby of kinetic isotope effects,[51,52] so applications to enzymic reactions began to appear. At first, even for modelling of non-enzymic reactions, it was often necessary to use a truncated version of the molecular system in order to reduce the number of atoms (and electrons) in the calculations. Certainly it was not possible to include any atoms of the protein environment in a model KIE calculation for an enzymic reaction, meaning that the cut-off approximation was being used implicitly. This fact also made the distinction between calculations for

enzymic or non-enzymic reactions rather meaningless! Nonetheless, arguably the first paper to report KIE calculations for a QM model of an enzymic reaction would appear to be the MNDO study by Ford and co-workers on cytochrome P450 oxidations of aliphatic substrates,[53] which was followed by similar studies of H-atom abstraction by Loew and co-workers for other substrates.[54]

These calculations used the Bigeleisen equation (Equation 19) with the rigid-rotor-harmonic-oscillator approximation and neglected tunnelling effects. However, some time ago it was pointed out[55] that there was some practical merit in separately determining the VP and MI terms for a pair of isotopologues (Equation 16), rather than relying upon their equality, particularly when the vibrational frequencies had been obtained using Hessians in Cartesian coordinates from standard packages for *ab initio* or semi-empirical molecular orbital calculations. Older implementations involving numerical second derivatives tended to produce Hessians which, upon mass-weighting and diagonalisation,[25] yielded eigenvectors corresponding to translation and rotation of the whole molecule which did not have zero eigenvalues: this meant that the Teller-Redlich product rule was not exactly satisfied, and consequently that application of the Bigeleisen equation incurred some error. Model calculations of KIEs for enzymic methyl transfer[56] (*ab initio* Hartree-Fock, HF) and carbonyl reduction by lactate dehydrogenase[57] (semi-empirical AM1) both explicit evaluation of the partition functions for translation, rotation and vibration after a projection to remove spurious translational and rotational components from the Hessian and to obtain six properly-zero eigenvalues.[55]

1.4.2 Dehydrogenases

An interest in KIE calculations for enzymic hydride-transfer reactions was initiated by the work of Jones and Urbauer on the identity hydride transfer from formate anion to carbon dioxide using a range of semi-empirical and ab initio HF methods.[58] No solvent was included because it could not be evaluated at the time, and the authors commented that 'since formate dehydrogenase is believed to act by desolvation of the substrate the gas phase should be a good model of the reaction in the enzyme active site provided specific amino acid interactions in the active site are small.' In making comparison between experimental (V_{max}/K_m) and calculated 1° ^{2}H, ^{13}C and ^{18}O KIEs (including a

Wigner tunnelling correction) for the formate moiety, they did not include the mass component of the MMI factor, but noted that both this and the inertial component (i.e. the translational and the rotational partition-functions contributions) might be affected by binding in a Michaelis complex. This was a prescient observation but it was not possible then to provide better estimates for the isotopic sensitivities of entropy changes in condensed-phase systems. Finally, Jones and Urbauer noted that a QM calculation yielded only a single KIE for isotopic substitution in a reaction whereas BEBOVIB was capable of allowing TS variations to be explored. Nonetheless they did suggest that the QM approach 'has the advantage of readily providing quality force fields, especially for larger systems, which otherwise would have to be fabricated from standard values or guesswork and trial and error.'[58]

The QM cluster approach was used by Andrés, Moliner and Safont to estimate 1° and 2° ^{2}H KIEs for the hydride-transfer step in lactate dehydrogenase (LDH), which catalyses the reduction of pyruvate to lactate.[59] Their model included the pyruvate substrate, the nicotinamide ring of the hydride-donor NADH cofactor, the imidazolium ring of the proton-donor histidine residue, and the guanidinium moiety of the carboxylate-binding arginine residue in the active site, a total of 44 atoms. Despite the lack of any further components of the protein or solvent environment, the TS structures located using several semi-empirical MO methods were consistent with what was known about other dehydrogenases. The RS structures were located by following the intrinsic reaction coordinate down from each TS to the adjacent energy minimum. The semi-classical KIEs (excluding tunnelling) for substitution of either the transferring (1°) or the non-transferring (2°) hydrogen at the C4 position of the nicotinamide ring were of plausible magnitudes as compared with experimental KIEs for another dehydrogenase (glutamate dehydrogenase). Note that since the calculated RS for the reaction was a model for the Michaelis complex within the enzyme active site, the calculated values presumably corresponded to KIEs on V_{max}. Hydride transfer is not the rate-determining step for the reaction catalysed by wild-type LDH, so the intrinsic KIEs are not immediately accessible experimentally.

Andrés and co-workers[60] applied a similar approach to calculate 1° ^{2}H and ^{3}H KIEs for hydride transfer from NADH to flavin adenine dinucleotide (FAD), using AM1 with a 45-atom truncated model and HF/6-31G for a 23-atom model. The equation they seem to

have used to calculate the semiclassical KIE appears to include only the ZPE factor and neglects the ratio of TS imaginary frequencies. Reasonable agreement with experimental isotope effects on V_{max}/K_m was apparently obtained but, as before, the RS in the calculations refers to the active-site complex immediately preceding the TS.

1.4.3 Binding isotope effects and software

Oxamate is an inhibitor of LDH, and its ^{18}O equilibrium isotope effect (EIE) for binding to LDH from aqueous solution was measured experimentally by Paneth and co-workers and computed by a variety of QM cluster methods.[61] This very small inverse EIE (0.9840 ± 0.0027) reflects formation of a hydrogen bond to carboxylate O atom in the enzyme active site. Using force constants computed for "full" models of oxamate in explicit water (22 molecules within a 5 Å radius) and oxamate in LDH (132 atoms including the sidechains of four amino acid residues and the nicotinamide moiety of NADH besides oxamate) with the AM1 semi-empirical method gave a calculated EIE of 0.9860 in very good agreement with experiment. However, using force constants computed for "cut" models of oxamate only within the frozen environment of the active site or solvent resulted in the calculated EIE being very much closer to unity (0.9988). This indicated the necessity of including more atoms in the Hessian evaluation than those of the inhibitor (or substrate) alone. Using some continuum solvation models to describe the aqueous solvation of oxamate gave good results in conjunction with the full model for the enzyme active site, but others did not, and the PM3 semi-empirical method performed badly by predicting the wrong direction for the EIE.

These calculations employed a version of Paneth's ISOEFF program[61] which enabled structures and Hessians to be input directly from several quantum chemistry packages (MOPAC/AMPAC, AMSOL, SIBIQ, GAMESS and GAUSSIAN) and computed KIEs by means of the Bigeleisen equation (Equation 19) and EIEs by a simple modification involving replacement of the TS terms by "PS" (product-state) terms involving $3N - 6$ frequencies, just as for RS. (The CAMVIB/CAMISO programs[29] of IHW, performed essentially the same function.) An upgraded program, ISOEFF98,[17] served to fill the gap between BEBOVIB and QM cluster calculations such that the user could modify selected QM force constants to see how these modifications affected the value of an isotope effect. First, the Hessian in Cartesian coordinates was transformed into (non-redundant) internal coordinates, then diagonal elements of the resulting force-constant matrix were

changed by hand while leaving the off-diagonal elements unchanged, and finally the back-transformation to Cartesians was performed before obtaining new vibrational frequencies from the modified Hessian. This new functionality was described by Anisimov and Paneth as "quantum-BEBOVIB".[17] The other widely used program for KIE calculations, QUIVER, described by Saunders, Laidig and Wolfsberg,[16] took Cartesian force-constant input from GAUSSIAN and computed the reduced isotopic partition function ratio for a pair of isotopologues by means of the Bigeleisen equation.

1.5 QM/MM calculations of KIEs

1.5.1 Early examples

Possibly the first example of a hybrid QM/MM method being applied to calculate KIEs was the study of alternative stepwise $D_N{*}A_N$ pre-association and concerted A_ND_N mechanisms for acid-catalysed AMP hydrolysis.[62] A 23-atom model for AMP and a nucleophilic water were described by AM1, and were surrounded by a sphere of about 500 TIP3P water molecules. A 2D grid search with respect to the bond distances from the anomeric C atom to the nucleofuge N and nucleophile O atoms generated an energy surface showing both pathways. Using a Hessian computed for the QM atoms only, the calculated 1° (anomeric) α-^{14}C and nucleofuge ^{15}N, and 2° β-^{2}H KIEs obtained from TS structures for both mechanisms agreed with each other and with the experimental values and could not serve to discriminate between the mechanisms. However, the calculated value for α-^{3}H was in excellent agreement with experiment for the stepwise TS but not the concerted TS. This result was in accord with Schramm's BEBOVIB-based interpretation.

The first application to an enzymic reaction was probably the AM1/CHARMM study of *p*-nitrophenyl-*N*-acetylneuraminide hydrolysis catalysed by influenza B and *Salmonella typhimurium* sialidases.[62] This work employed a QM region of 65 atoms (including a nucleophilic water, truncated substrate and sidechains of active-site aspartate, glutamate and tyrosine residues) and an MM region of about 1500 surrounding atoms within a stochastic boundary. Approximate RS and TS structures were located by a 2D grid search (as for AMP hydrolysis) and 1° nucleofuge ^{18}O and 2° β-$^{2}H_{(R)}$, β-$^{2}H_{(S)}$ and β-$^{2}H_2$

KIEs were computed. It was commented that comparison of calculated KIEs with experiment provided a very stern test, and their use provided 'a very strong anchor to prevent theoretical modelling from drifting into unreality'. The less-than-wholly-impressive agreement of calculated and experimental KIEs (for influenza A sialidase) led Barnes and Williams to suggest that their QM/MM modelling 'had not yet passed this test' and to speculate that the procedure employed for optimising and characterising stationary points might have been at fault.

1.5.2 Hydride and hydron transfer

A more rigorous procedure was developed as a suite of utilities (GRACE) for optimisation of minima and saddle points, IRC paths and vibrational frequencies in large flexible QM/MM systems, which enabled TSs to be refined using the same criteria as were normally used in QM applications with small molecules, and this methodology was applied to LDH.[63] A Hessian was computed for a "core" of 147 atoms comprising 35 QM (AM1) atoms and 112 MM (CHARMM) atoms with in the active site; the QM region was embedded mechanically and electronically within the MM "environment" of about 1900 atoms, and was a subset of the core atoms included in the Hessian. Although the residual force on every core atom was tolerably close to zero within the full QM/MM system, the core was not itself a stationary structure in isolation. Transformation from Cartesians to non-redundant valence coordinates, followed by back-transformation, served to project out the components of translation and rotation of the core as a rigid entity with respect it its environment,[28] equivalent to treating the core atoms as gas-phase cluster with force constants determined within the protein environment. The average value of the 1° ^{2}H for hydride transfer from NADH to pyruvate, taken over six different TS configurations within the flexible enzyme active site, was 2.73 ± 0.06 at 298 K, which compared well with experimental values for mutant LDHs with rate-limiting chemistry as opposed to a conformational change as in the wild-type).[64] Moliner and co-workers subsequently applied the same method to the hydride transfer step in dihydrofolate reductase.[65]

The proton transfer reaction involved in conversion of 2-phospho-d-glycerate to phosphoenolpyruvate catalysed by yeast enolase was studied by Truhlar and co-workers[66] using a AM1/CHARMM hybrid method with a 25-atom QM region surrounded

by 8863 MM atoms. Their calculation of the semiclassical 1° ^{2}H KIE was very similar to that used for LDH,[63] except that the generalised hybrid orbital method was used to define the QM/MM boundary instead of the link-atom method; the (conventional) TS Hessian was determined for the N_1 = 25 atoms QM atoms at the saddle point in what they termed as the primary zone, with vibrational partition functions being evaluated for $3N_1 - 1$ degrees of freedom. However, quantum effects along the reaction coordinate were treated by the microcanonical optimised multidimensional tunnelling approximation with small-curvature tunnelling (SCT), which also includes non-classical reflection by the reaction barrier: this gave a KIE of 3.5 at 300 K as compared to the experimental value of 3.3. A separate determination of the quantal corrections within the primary zone was required for each isotopic substitution.

A similar treatment of hydride transfer from a benzyl alcoholate substrate to NAD^+ catalysed by liver alcohol dehydrogenase (LADH) was also reported by Truhlar, Gao and co-workers, but ^{2}H and ^{3}H 1° and 2° KIEs were determined as averages over 20 configurations of the 21-atom (mobile) primary zone within the 5539-atom (frozen) secondary zone.[66] The merits of the multidimensional treatment of tunnelling were clearly stated in comparison with the one-dimensional Bell correction applied by Rucker and Klinman in a BEBOVIB analysis of a truncated model for LADH.[67] In order to obtain satisfactory agreement between calculation and experiment for the 2° ^{3}H KIE in the 1D parabolic method, the tunnelling correction was overestimated, whereas the 63-dimensional non-parabolic method reproduces this KIE and the Swain-Schaad exponent in agreement with experiment. Another paper on LADH by the same group made further refinements to the method and noted that the quantum effects in the primary zone were included 'only on the $3N_1 - 7$ highest-frequency vibrational modes'.[68] The transmission coefficient was now evaluated within a relaxed secondary zone, and the procedure allowed for Boltzmann averaging over an ensemble of RS and TS conformations.

Cui, Elstner and Karplus used related methodology, based upon a QM/MM method using the self-consistent-charge density-functional-tight-binding (SCC-DFTB) method with CHARMM, with a much larger number of mobile atoms (278) in the secondary zone but with no averaging over different enzyme structures;[69] these authors obtained similar KIE results to those of Gao, Truhlar and co-workers.[68] In contrast, Tuñón, Moliner and

co-workers noted that reactions taking place within separate valleys, with distinctly different reactant conformations with the active site of LDH, gave different 1° ^{2}H KIEs as evaluated a slightly modified version of the ensemble-averaged variational TS theory (EA-VTST) with multidimensional tunnelling contributions.[70] Burton and co-workers also used a similar approach to calculate 1° ^{2}H KIEs for proton transfer from a methylamine substrate catalysed by methylamine dehydrogenase (MADH) for a range of enzyme conformations; large tunnelling corrections were predicted for proton transfer with both semi-empirical and ab initio QM/MM methods.[71] Ranaghan *et al.* performed similar calculations for MADH, using PM3 (with specific reaction parameters) and CHARMM together with EA-VTST/SCT, and averaged their 1° ^{2}H KIEs over different minimum-energy paths. They found that the protein configuration used had a significant effect on the computed KIE, with average values of 14 or 41 being obtained for proton transfer to one of the O atoms of an aspartate carboxylate sidechain for two ensembles of structures taken from different regions of the reaction coordinate.[72] These authors also applied the same methodology to calculate KIEs for oxidation of tryptamine by aromatic amine dehydrogenase.[73]

The hydrogen-atom transfer catalyzed by coenzyme B_{12}-dependent methylmalonyl-CoA mutase was studied by Dybala-Defratyka *et al.* using AM1/CHARMM and EA-VTST with averaging over eight reaction valleys descending from points selected from the TS ensemble. The importance of corner-cutting tunnelling paths in order to obtain a 1° ^{2}H KIE that agreed with experiment within experimental error was demonstrated.[74] The 45-atom QM region was described by an unrestricted Hartree-Fock implementation of the semi-empirical MO method within the CHARMM program, and the dynamics calculations employed CHARMMRATE[66] with a 45-atom primary zone. Agreement with experiment could not be obtained from smaller QM cluster models that included part or all of the substrate and co-factor moieties but excluded active-site protein residues.[75] Without the enzyme, a larger KIE of about 100 suggested the existence of a higher commitment to catalysis, and consequently led to the conclusion that the intrinsic KIE was masked by the kinetic complexity, whereas with enzyme residues included in the QM/MM model, the good agreement between calculation and experiment was consistent with a minimal commitment.

1.5.3 Chorismate mutase

Chorismate mutase has been to QM/MM studies of enzyme catalysis in the 2000s what the water molecule was to electronic structure theory in the 1960s: a very convenient benchmark. Since it catalyses the pericyclic rearrangement of chorismate to prephenate, including these substrate atoms in the QM region, while treating the whole protein within the MM region, means that the QM/MM boundary does not cut through any covalent bonds. A 2003 study by Martí *et al.*[76] used an AM1/CHARMM24/TIP3P hybrid method to locate any characterise RS and TS structures with full gradient relaxation of > 5200 atoms (out of a total of > 17000 MM atoms) for the enzymic simulations, and with a box containing 711 water molecules for the corresponding reaction in aqueous solution. Inclusion of the chorismate conformational pre-equilibrium step in the modelled kinetic scheme led to better agreement between calculated and experimental 1° ^{13}C and ^{18}O KIEs. Good agreement for the 2° **5-**^{3}H and 9-$^{3}H_2$ KIEs in aqueous solution was obtained provided that the Hessian water molecules hydrogen-bonded to the substrate, but poorer agreement for the enzymic reaction may have been due to the available experimental V_{max}/K_m KIEs being for chorismate mutase from a different organism to that for the available X-ray structure used as the starting point for the calculations. Roitberg and co-workers calculated KIEs for the rearrangement step, but not the conformational pre-equilibrium, using an ab initio HF treatment of the substrate with effective fragment potentials to represent the protein environment.[77] The same system was later studied using "high-level" (DFT) corrections were applied to the "low-level" (AM1) QM subsystem during QM/MM evaluations of gradients and Hessians.[78] This procedure arguably gave more reliable estimates of the intrinsic KIEs for the rearrangement, and discrepancies with experiment may have been due to the rearrangement occurring after the rate-limiting step for the enzymic reaction.

Williams and co-workers[79] used their original AM1/CHARMM24/TIP3P hybrid method to investigate the sensitivity of calculated KIEs to the size of the Hessian for the chorismate rearrangement in water and in the active site of chorismate mutase. As the system was expanded to include not only the atoms of chorismate/prephenate itself but also an increasing number of surrounding water molecules (up to 99) or active-site residues (up to 225 atoms), 1° ^{13}C and ^{18}O KIEs were not sensitive to the size of the Hessian, but 2° 5-^{3}H and 9-$^{3}H_2$ KIEs required the inclusion of at least those atoms

directly involved in hydrogen bonds to the substrate or, better, a complete first solvation shell or cage of active-site amino-acid residues.

1.5.4 Methyl transfer and the compression hypothesis

Few species in enzymology have sparked such fervent controversy as catechol-*o*-methyltransferase, an enzyme which catalyses the transfer of a methyl group from *s*-adenosyl-methionine to dopamine [80] (Figure 3)

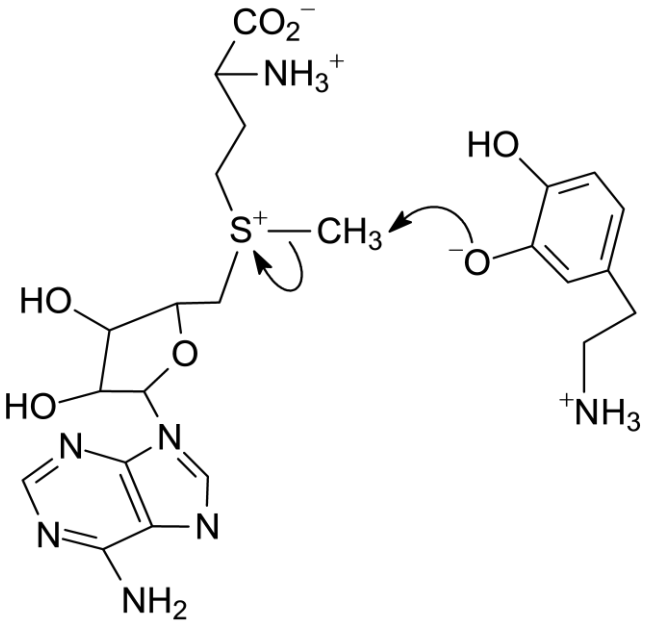


Figure 3. Diagram of proposed mechanism for methyl transfer from adenosyl moiety to catecholate. Proposed mechanism from N. Kanaan, J. J. R. Pernia and I. H. Williams, Chemical Communications, 2008, **46**, 6114-6116.[80]

Theories regarding the origins of enzyme catalysis are substantial in number. Quantum tunnelling has many adepts [81-83], as well as detractors [84, 85], one of the latter being Arieh Warshel, co-Laureate of the 2013 Nobel Prize in Chemistry. His research strongly suggests that tunnelling does not significantly contribute to enzyme catalysis, by comparison to the reactions in solution, and the enzyme active site. Conversely, other groups claim support of their results through evidence from large normal isotope effect measurements on methyltransferases [86]. These groups also claim so-called *dynamic* effects are responsible for enzyme catalysis. One of their leading hypotheses focuses on

a compression effect of the enzyme active site, which mechanically forces catalysis to occur. This is attributed to the S_N2 transition structures being subject to compression, leading to larger inverse 2^o KIEs for COMT-catalysed methyl transfer, compared to a reference reaction in water.[87]

The compression hypothesis for the origin of catalysis in COMT (and other group-transfer enzymes) has had both its supporters and detractors over the years since the pioneering BEBOVIB studies of Schowen and co-workers. A hybrid AM1/TIP3P/CHARMM calculation of 2° α-$^2H_3$ KIEs for the reaction of *S*-adenosylmethionine (AdoMet) with catecholate anion in water and catalyzed by COMT gave calculated KIEs in agreement with experiment, with a significantly more inverse KIE for the enzymic reaction than for the one in aqueous solution.[88] However, this result did not support the compression hypothesis because the sum of the bond lengths from the α-C atom to the S atom of the nucleofuge and to the O atom of the nucleophile was essentially the same in the COMT active site as it was in water; in other words, there was no evidence for geometrical compression. The point was further reinforced by a subsequent study in which the KIEs were evaluated as averages over RS and TS structures taken from QM/MM molecular dynamics (MD) trajectories.[89] The QM region comprised 63 atoms (catecholate + AdoMet) and the full QM/MM system for the MD simulations contained 17162 atoms including 4614 water molecules; the flexible region for local relaxation of "snapshot" structures (to either a local minimum or saddle point) contained 2610 atoms; and the Hessian included the QM atoms and the active-site Mg^{2+} cation (64 atoms). The uncatalysed methyl transfer in solution was the same as that studied experimentally in the original work by Schowen's group, namely reaction of methoxide anion with *S*-methyldibenzothiophenium cation in methanol; the QM region comprised 30 atoms and the Hessian included at least a complete first solvation shell of solvent molecules around the solute. MD trajectories of 1 ns duration were performed in RS and TS windows from a potential of mean force (PMF) along the reaction coordinate and 100 snapshots were taken at 10 ps intervals in order to ensure that all structures were independent, and all simulations were performed at the temperatures of the experimental studies. The mean value of the calculated KIE for the enzymic reaction, averaged over 10000 combinations of RS and TS (0.82 ± 0.05) agreed well with the experimental value (0.83 ± 0.05), but it was noted that the standard deviation for the calculated KIE reflected the uncertainty of averaging over precisely computed individual

TS-theoretical rate-constant ratios for a limited sample of 100 RS and 100 TS configurations spanning a distribution of structures determined by the model potential, whereas the experimental errors arose from uncertainties in rate-constant determinations for (effectively) complete sampling of all thermally accessible RS and TS configurations which may also span a distribution of structures. It was suggested that averaging over thermal fluctuations of the environment allowed for meaningful comparison of computed with experimental KIE values. Furthermore, the authors commented that apparent differences between KIEs calculated by different QM methods were often smaller than the fluctuations found between different configurations within the same method, with the implication that computational studies based upon a single conformation of an enzyme might be unrepresentative.[80]

Analysis of valence force constants computed in the 2004 QM/MM study for the COMT-catalysed methyl transfer showed that the methyl CH stretching force constants were larger in TS than in RS, as expected.[90] What was not expected, however, was that the HCS and HCO bending force constants, involving these bonds and either the nucleofuge S atom or the nucleophile O atom, were also significantly larger in TS than in RS, both in the COMT active site and in solution. These comparisons were made using relaxed force constants (reciprocals of diagonal elements of the compliance matrix in valence coordinates) which have the property of being independent of the choice of redundant internal coordinates.[29] This finding shed additional light on the compression hypothesis. The TS bending force constants used in the original BEBOVIB modelling were obtained by use of a rule that took the product of the RS force constant (for HCS) or the PS force constant (for HCO) and the appropriate TS bond order (B_{CS} or B_{CO}) which always had a value < 1; the TS bending force constants were, therefore, always less than the corresponding RS (or PS) value. In order to calculate an inverse 2° α-$^2H_3$ KIE to match that of the enzymic reaction, it was necessary to use larger values for B_{CS} or B_{CO} than for the uncatalysed reaction. The apparently shorter bond lengths for the C…S and C…O partial bonds in the TS were probably an artefact of the arbitrary rule used to generate TS bending force constants within BEBOVIB.

In 2011, Klinman and coworkers published a study on active site residues and their roles in generating active site compaction in COMT-catalysed methyl transfer. Comparing catalytic efficiency with the size of a 2^o KIE, Klinman found a linear

correlation, interpreting this on the basis of modulation by a side chain located near the sulphur of the AdoMet.[91] Additional work by the group, published in 2015, suggested that increases in C-H force constant in the methyl group at the active site were as a direct result of compaction in the donor-acceptor distances.[92] Previous work by other groups had suggested that no such compaction occurred, by comparison of the donor-acceptor distances themselves,[56, 90, 93] yet the question remained as to which additional factors could contribute to this change in force constant, and in what manner could these be quantified.

1.5.5 Other enzymes

Kwieciek and co-workers[94] studied the cytochrome-P450 2A6-catalysed hydroxylation of nicotine, at either of two positions, using B3LYP with the 6-31G(d) basis (LACVP for Fe) in combination with the AMBER force field in an ONIOM method with mechanical embedding. The approximately 70-atom QM region consisted of the truncated heme, cysteine and nicotine, and about 1600 MM atoms of the active site were included. KIEs for six different isotopic substitutions were computed for both low-spin and high-spin states of iron using ISOEFF98. Comparison of the KIEs predicted for reaction at each of the two sites gave mechanistic insight.

A QM/QM ONIOM method, with 30 – 40 atoms in the higher layer (B3LYP/6-31G(d)) and 800 – 900 atoms in the lower layer (PM6), was applied by Paneth and co-workers to calculate 1° ^{37}Cl KIEs on the nucleofuge in hydrolysis of both enantiomers of 2-chloropropionate by DL-2-haloacid dehalogenase.[95] Excellent agreement with experiment was obtained provided that account was taken not only of the KIE on the S_N2 step but also of the EIEs for binding of the substrate (first) from water to an initial binding site and (second) from there to the active site. Krzemińska *et al.* performed QM/MM calculations of heavy-atoms (^{13}C, ^{15}N, and ^{18}O) EIEs for binding sites of different inhibitors to HIV-1 reverse transcriptase.[96] The QM region (either AM1 or B3LYP/6-31+G(d,p)) consisted of the inhibitor molecule alone and the MM region (AMBER for protein or TIP3P for aqueous solution) comprised the rest of the system. The subset

Hessian was projected to remove the translational and rotational contributions, and the binding isotope effects (BIEs) were obtained as the average over 121 pairwise combinations of 11 RS (solution) and 11 PS (enzyme-bound) structures. Significant BIEs (especially ^{18}O) were found for binding of nucleoside HIV-1 RT inhibitors by means of specific interactions in a hydrophilic binding site, whereas no BIEs were found for non-nucleoside HIV-1 RT inhibitors binding within a hydrophobic cavity on the protein. Essentially the same method was used by Świderek *et al.* to compute KIEs for hydride transfer from tetrahydrofolate to the exocyclic methylene–dUMP intermediate in the reaction catalysed by thymidylate synthase.[97] QM/MM free-energy surfaces were determined by means both 1D and 2D PMFs which allowed concerted and stepwise reaction paths to be characterised. The 54-atom QM region was described by AM1 and by a dual-level method involving M06-2X corrections; the rest of the system was described using the OPLS-AA and TIP3P force fields. The AM1/OPLS-AA KIEs were averaged over all combinations of 10 RS and 10 TS structures. The 2° ^{3}H KIE for substitution at C6 of tetrahydrofolate (from which hydride transfer occurs) was revealing: a small normal effect, consistent with experiment, was calculated for the concerted mechanism, but an inverse effect was predicted for the alternative stepwise mechanism. The calculated KIEs served to discriminate between the two mechanisms.

1.6 Summary

This chapter has concentrated on the theory of isotope effects, and important developments in the field over the past century. Kinetic and Equilibrium isotope effects are used as probes for a range of chemical systems, and their reliable calculation and interpretation is vital in order to successfully treat the information gathered.

Developing on the theme of theory and setting; the following chapter will treat the core computational and theoretical concepts and methodologies applied during the course of the work presented.

2. Computational Theory, Experimental Details and Isotope Effect Calculations

2.1 Introduction

Ever since the availability of computers was extended to general scientific audiences, they have been used to solve complex theoretical problems [98]. Quantum chemistry has the advantage that, compared to certain fields in physics, which are still difficult to explain with contemporary science, it is possible to gain reasonable quantitative answers to questions about molecules [99]. This is mainly based on the wide range of problems that the solution to the Schrodinger equation can be applied to, largely yielding answers of good accuracy to experiment [100].

2.2 Density functional theory and QM calculations

A number of approximations are responsible for our ability to predict molecular properties. The Born-Oppenehimer approximation [101] allows the motion of electrons only to be considered, as the massive nuclei are taken as moving too slowly on the same timescale to affect results with any significance, and is the cornerstone of modern QM chemical approaches.

Developing on this, Kohn, Sham and Hohenberg formulated a way to simplify the complex many-body wavefunction for a system, into a density which depends only on three coordinates [102, 103]. Density Functional Theory (DFT) is based on the use of *functionals*, which provide a reasonable balance between accuracy and computational expense. The incorporation of these functions of functions provide an approximate treatment of electron correlation – the effect one electron has on another [104]. This is not only absent from the earlier Hartree-Fock methodology, but is often exceedingly computationally expensive in higher order methods such as perturbation (MP2…) and coupled cluster (CCSD…) [105,106].

Density functionals form the backbone of DFT and their parameterisation determines the accuracy of the properties obtained. Functionals are often ranked in order of sophistication, from the Local Density Approximation (LDA) and Generalised Gradient

Approximation (GGA), to hybrid methods [107]. However, these postulated improvements are no reliable indication that results will be improved. This is contrasted by *ab initio* methods, which relatively always provide superior computation with a higher method. Using a coupled cluster method will almost always be superior than the same calculation with MP2 and that in turn more accurate than simple Hartree-Fock [108].

Density functionals constitute of a number of terms in order to successfully describe the density and reasonably calculate molecular properties. The kinetic energy, electron-nuclei interaction, Coulomb interaction and exchange-correlation energy all form part of functionals [109].

Fortunately, the kinetic energy, electron –nuclei interaction and Coulomb energy can all be computed exactly [110]. The difficulty is incorporating the exchange-correlation energy. Due to the complex behaviour of electrons, this would be inherently difficult to calculate exactly, which led to numerous approximate exchange-correlation functionals being developed [111].

These are often described in a hierarchy based on their parametrisation (Table 1). LDA-based functionals are frequently considered the simplest implementations and are based on the electron density at a given point, using a uniform electron gas as a model. LDA functionals were improved by including the gradient of the density as well as its local behaviour. This means that the variation of this key parameter is taken into account when implemented. Realistically, few systems behave as uniform electron gases. It is more usual to find the electron density varying significantly over a small region, thereby suggesting that a uniform electron gas as a model may lead to erroneous results.

Table 1. Selected functionals arranged in a Jacob's ladder classification of approximations to the exchange correlation energy, from the most fundamental early methods such as LDA, to methods of higher accuracy at PBE0 level.[112]

Level	Name	Examples
1	LDA	LDA, LSDA [113]
2	GGA	BLYP, PW91[104], PBE [114]
3	Meta-GGA	BR, B95 [115], VSXC
4	Hyper-GGA	B3LYP, PBE0 [116]

GGA functionals were themselves improved by including the derivative of the gradient, or second derivative, and are found in examples such as M06-L, classed as a Meta-GGA functional.

Hybrid functionals are currently the improvement on Meta-GGA as include Hartree-Fock exchange in the parameterisation. B3LYP is a common hybrid functional which is known as Hyper-GGA as mixes a fixed amount of HF exchange with GGA exchange and local hybrids.[117] B3LYP performs well for numerous systems. Hybrid functionals were first discussed by Becke in the early 1990s; they consist of the traditional formulation, with the exchange-correlation energy derived from an interpolation between the real system, and an entirely non-interacting system [98].

Therefore, in order to run a DFT calculation, the molecular geometry and charge must be specified, as well as the multiplicity of the system. In terms of the parameterisation, the exchange and correlation functionals must be included, as well as a basis set [118].

Alternatives to DFT exits, some referred to previously in the text. Semi empirical methods simplify the way in which the wavefunction is obtained by incorporating experimental data into the procedure. This results in a faster technique which works well for certain sets of molecules, but can result in large errors for systems deviating from the experimental data range.

2.2.1 Dispersion effects in QM calculations

London had originally described dispersion in an additive sense; four pairwise interactions would sum to yield the energetic contribution to stabilisation from dispersion. In more recent studies, it has been found that van der Waals interactions are not additive, and indeed more complex.[119] The interactions between molecules is affected by their interactions with the environment, and surrounding molecules; a summation of pairwise interactions is therefore not sufficient to describe the complete interaction energy of a molecule with those surrounding it. Two principle methods to account for London dispersion have been formulated. The first, symmetry-adapted perturbation theory, or SAPT[120] is based on predefined molecular moieties, where the

interaction energy is calculated by small perturbations to the system. The total energy is calculated from individual parts of the molecular complex, and is therefore physically meaningful. Difficulties arise when the interactions between molecular moieties are large, and can therefore not be treated through perturbation approaches. The alternative method, the supramolecular approach, calculated the interaction energy as the difference between the total energy and those of the monomers. Understandably, this method has difficulties with covalent complexes. Indeed, for dispersion effects in chemical reactions, it is suggested that two levels of theory be implemented, one incorporating dispersion, and the other without.

Stefan Grimme has proposed empirical dispersion corrections for density functional calculations,[114] and these are incorporated into common electronic structure calculations software.

The total energy of a system with dispersion corrections can be described as:[121]

$$E_{DFT+D} = E_{DFT} + E_{dispersion} \tag{20.1}$$

Where E_{DFT+D} represents the total energy of the system with dispersion correction implemented, E_{DFT} is the traditional Kohn-Sham DFT energy output from using the chosen functional, and $E_{dispersion}$ is the energy contribution from the dispersion interactions.

The energy contribution from dispersion can be described by:[122]

$$E_{dispersion} = -s \sum_{i=1}^{N_{atoms}-1} \sum_{j=i+1}^{N_{atoms}} \frac{C_6^{ij}}{R_{ij}^6} f_{damping}(R_{ij}) \tag{20.2}$$

Where s is a scaling factor based on the density functional chosen for calculation or the electronic structure and properties, N_{atoms} is the number of atoms in the system, C_6^{ij} is the dispersion coefficient for the pair of atoms i,j and R_{ij} is the distance between atoms

i and *j*. $f_{damping}$ acts as a damping function for when R_{ij} approaches 0, and would alternatively lead to artificial overbinding of *i* and *j*. Dispersion corrections can be implemented within calculation routines in a number of ways. Dispersion can be included in the optimisation process (DO), and lead to dispersion-corrected structures, or can alternatively simple be called for single point energy corrections (DC). These themes will be discussed further in later sections of this thesis.

2.3 Basis sets

Basis sets can be considered as the following level of calculation accuracy after having carefully chosen the electronic structure method (DFT, MP2, CCSD…). In order to treat the electrons sufficiently well, they must be allowed to occupy a certain volume in space. This is defined by the basis set; a series of formulae which describe the space afforded for the electrons within a molecular simulation [123].

Basis sets vary in nomenclature and parameterisation. Commonly-used basis sets in molecular modelling include the so-called *Gaussian* family of basis sets, where electrons are described as Gaussian functions. Although less reasonable representations of the behaviour of atomic orbitals *cf.* Slater type orbitals, Gaussian basis functions are preferred as are more facile to combine in computations through simpler formulae [124]. Gaussian basis functions can form the commonly-used *Split-valence* basis sets, which include 3-21G, 6-31G(d), and 6-311+G(d,p) [125] amongst others. Split-valence basis sets are commonly used in DFT and post-HF calculations. These descriptions of electrons rely on two functions: an inner, and an outer. This results in core atomic orbitals (AOs) being characterised by one set of functions, and valence AOs by another. The 6-31G basis set, used in some of our work, consists of 6 basis functions for each core orbital, 3 for inner valence orbital components, and one for outer valence orbital components. This splitting of the valence orbitals is advantageous when describing different characteristics of the same orbitals. For example, a split valence basis set would cope well with defining both sigma- and pi- type p orbitals, both of which behave differently, the pi- creating a weaker/looser bond than the sigma.

Polarisation and diffuse functions [126] can be added to basis sets in order to provide additional descriptions for electron motion. Polarisation functions add higher-orbital functions *i.e.* adding d-functionality to elements with s and p valence orbitals (indicated

by a '*' in the basis set description). Diffuse functions act as the name suggests – allowing the electrons to be only weakly associated with the atom. Normally, diffuse functions are only added to atoms larger than hydrogen (denoted '+'), as these will benefit more from the enhanced electronic freedom, although diffuse functions for hydrogen ('++') can also be added.

A range of electronic structure methods and basis sets have been surveyed throughout the course of this work, and in order to avoid repetition, their relative merits and issues will be considered as they appear in the course of the narrative. Evidently, these method validation tasks tend to be respective to the system studied. It is unlikely that a particular combination of ESM and basis set will provide as particularly reliable results for one enzymic reaction, as it would for another. Therefore, a series of method validation exercises took place before the introductory studies into the nature of isotope effects.

2.4 Solvation methods in QM calculations and Enzymology

As discussed in Chapter 1, increases in computing power have brought with them additional functionality to chemical calculations. Early QM calculations were limited to small-molecule, gas phase systems, regardless the influence of solvation on the system.[127] Solvation can be incorporated into a chemical calculation in numerous ways, most of which can be characterised as either *explicit* or *implicit*.

Explicit solvation involves describing every solvent molecule individually. Practically, this is carried out through Monte-Carlo (MC) or molecular dynamics (MD) simulations of a box containing a fixed number of molecules of solvent. Here, all types of interactions are accounted for; however the computational cost is relatively significant. Rather than describe all solvent molecules individually with the same ESM as the central body of the system, a cutoff is often used to consider only the first series of solvation shells around the reaction centre with the higher level methodology. Subsequent shells are then described with a lower level, such as an MM force field or equivalent. This provides the effect of the outer forces and charges on the core region, without the significant expense of high-level methods. The choice of force field is therefore important, and numerous

parameterised instances exist. These tend to be generated for a specific molecule of set of molecules, which are detailed in the brief for the potential.

In order to reduce the expense, implicit solvation models were developed. These rely on placing the solute in an average field (constant dielectric), appearing in the calculations as an extra parameter akin to HF calculations and the core Hamiltonian. The main disadvantage of implicit models is the inability to account for interactions such as hydrogen bonds, and therefore have difficulty describing such systems accurately.

One instance of implicit solvation is the polarised continuum model (PCM) developed by Tomasi and co-workers.[128] Here, the solute is placed in a cavity surrounded by a volume of average dielectric, which must be specified. The shape of this cavity is determined by the United Atom Topological (UA0), or Universal Force Field (UFF) models. UA0 places a sphere around heavy atoms (larger than hydrogen), with hydrogens being encompassed within the heavy atom sphere.

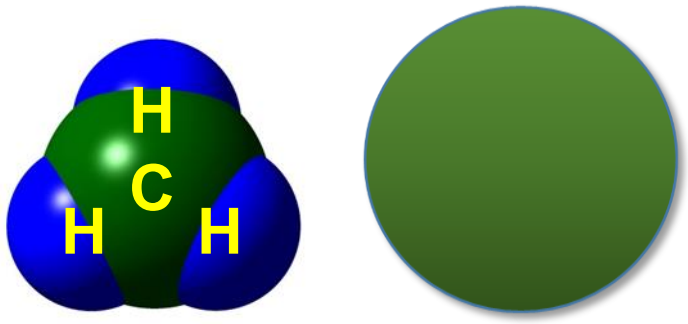


Figure 4. Representations of the UFF (left) and UA0 (right) cavities, respectively.

Conversely, the UFF cavity consists of spheres placed around each atom, proportional to their size, resulting in greater exposure of the system to the dielectric. This is the subject of Chapter 3, and will therefore be discussed in additional detail there.

In summary, solvation in chemical calculations can be described by a number of methods, each with subsets of changeable parameters. The importance remains with considering the options available to the calculation and choosing the more reliable

model for the study itself. For example, larger systems would likely benefit more from a hybrid approach, such as QM/MM systems, which are treated in section 2.5.

2.5 QM/MM methodology

Having discussed implicit solvation in section 2.4, it is timely to explain the method by which minimised supramolecular structures have been obtained for the calculation of isotope effects. MM techniques by themselves represent molecular interactions in a simpler way than QM approaches, and are therefore less suited to examining certain processes such as chemical reactions. The benefits of implicit solvation for smaller systems have been discussed, and the reliable chemical data that can be obtained from *ab initio* and QM approaches, however for systems with a more significant number of molecules, a partition between QM and MM methodologies is beneficial. This allows the central reaction volume to be treated by the higher level QM technique, whilst the surrounding is modelled with an appropriate MM force field, still inducing an effect on the reaction centre, particularly in terms of charge.

The first QM/MM potentials were developed in the 1970s by Warshel and Levitt, [129-133] being used to study the damage to bacterial cell walls by the enzyme lysozyme and the mechanism of the chemical reaction it catalyses. Due to the size of enzymes, it was indeed too large a number of atoms to study with a quantum chemical methodology, which led to Warshel and Levitt restricting the QM calculation to the smaller number of atoms directly associated with the chemical reaction being studied. With QM methods being particularly limited by CPU power in the 1970s, a semiempirical method was chosen for the QM region, the rest of the system being treated by a MM force field. In 1990, Martin Karplus, along with colleagues Martin Field and Paul Bash produced a combined QM/MM potential with the MNDO and AM1 semiempirical methods, and the CHARMM forcefield as the MM portion.[134] They tested this against ab initio and experimental data available, in order to consider the accuracy of their proposed method. QM/MM methods and their applications to biological systems have been the subject of

numerous recent reviews,[135-150] with more modern techniques such as path integral methods featuring heavily.[151] There are a vast number of QM/MM potentials available, but the same general theory applies to all approaches. An excellent summary of applications of QM/MM methods in recent biomolecular research, as well as a list of specific potentials and techniques applied was produced by Senn and Thiel.[152]

2.5.1 Theoretical basis of QM/MM calculations

Biological systems are by nature complex. It is therefore important to be able to treat these with reasonable accuracy and resources, and within reasonable timescales. Molecular dynamics (MD) simulations rely on solving Newton's equations of motion in order to obtain trajectories of molecular movement with time. QM methods must be used to model the electronic effects and their change over the course of a chemical reaction, however this does not need to be extended to all parts of the system, simply those involved in the reaction. Therefore, the atoms surrounding the central reacting region can be treated with less computationally-expensive techniques, such as potential models or molecular mechanics (MM) forcefields.

In MM forcefields, basic chemical theory is used to obtain the potential energy of the system.

$$V_{MM} = \sum_{i}^{N(bonds)} V_i^{bond} + \sum_{j}^{N(angles)} V_j^{angle} + \sum_{k}^{N(torsions)} V_k^{torsions} + \sum_{i}^{N(MM)} \sum_{j>i}^{N(bonds)} V_{ij}^{Coulomb} + \sum_{i}^{N(MM)} \sum_{j>i}^{N(bonds)} V_{ij}^{IJ} \quad (21)$$

In Equation 21, *N(MM)* refers to the number of atoms within the system, with V^{bond}, and V^{angle} being treated by harmonic potentials. Torsions are treated by periodic functions ($V^{torsion}$), and the Coulomb law is represented by $V^{coulomb}$, where R_{ij} is the distance between atoms *i* and *j*, *e* is the unit charge, and ε_0 is the permittivity of free space:

$$V_{ij}^{Coulomb} = \frac{e^2 Q_i Q_j}{4\pi\varepsilon_0 R_{ij}}$$

(22)

The additional parameter denoted V^{IJ} corresponds to the Van der Waals interactions, in the form of a Lennard-Jones potential:

$$V_{ij}^{IJ} = \left(\frac{A_{12}^{ij}}{R_{ij}}\right)^{12} - \left(\frac{B_6^{ij}}{R_{ij}}\right)^{6}$$

(23)

The *A*, and *B* constants in Equation 23 correspond to repulsion and attraction parameters, respectively, depending on the types of atom *i* and *j*. Electrons therefore do not feature in the parameterization of the MM forcefield, instead being accounted for by the empirical parameters present. Chemical reactions and other electronic processes can therefore not be described at the MM level of theory. Hybrid QM/MM methodologies therefore allow these reacting systems to be considered.[153]

Hybrid QM/MM potential energies include contributions from three main types of interaction. These are formed of interactions between QM atoms, interactions between atoms in the MM region, and interactions between the QM and MM atoms Two forms of scheme have been formulated to account for these interactions. The first, termed the *subtractive* coupling scheme has the QM/MM energy of the system obtained in three steps:

$$V_{QM/MM} = V_{MM}(MM + QM) + V_{QM}(QM) - V_{MM}(QM)$$

(24)

The final portion of the subtractive scheme effectively accounts for including the QM subsystem interactions twice. Subscripts indicate level of theory, and bracketed terms the atoms included in the term.

Contrary to the subtractive scheme, the additive scheme consists of a sum of the MM, QM and QM/MM energy terms, with the QM system embedded inside the large MM system:

$$V_{QM/MM} = V_{QM}(QM) + V_{MM}(MM) + V_{QM-MM}(QM + MM) \quad (25)$$

Only MM interactions are described by forcefields in the additive scheme; the MM interactions hold the QM subsystem in place.

2.5.2 ONIOM

In addition to traditional QM/MM schemes as detailed above, other approaches based on layered models such as ONIOM have been employed.[154-164] This approach is based on considering the molecules as two systems, a real system consisting of all the atoms in the simulation, and a model system made up of only the QM atoms. The energy of the real system can then be approximated as the sum of the model QM and real MM energies, subtracting the model MM energy. This has the result that the QM region energy is effectively calculated three times, once for each of the model system calculations, and once as part of the real system. In order to mediate this and only account for the QM system once, the model MM energy is subtracted. ONIOM is therefore described as a subtractive scheme, as opposed to additive methods discussed previously. Additionally, in more rudimentary ONIOM calculations, the QM interactions with other layers are only treated with an MM potential as the model system does not include the surrounding atoms. A simple ONIOM energy evaluation is described in Equation 26.[165]

$$E_{ONIOM} = E_{QM(m)} + E_{MM(r)} - E_{MM(m)} \quad (26)$$

Where $E_{QM(m)}$ is the QM energy of the model system, $E_{MM(r)}$ is the MM energy of the real system, and $E_{MM(m)}$ is the MM energy of the model system. This subtractive scheme

effectively removes the double counting in the MM contributions to the total ONIOM energy.

The ONIOM method is used in numerous investigations and both it and traditional hybrid QM/MM methodologies have their advantages and disadvantages, the main issues arising from the manner in which the simulation is partitioned into different sections. For a thorough review of ONIOM methods and associated techniques, see Morokuma *et al.*[165]

2.5.3 QM/MM partitioning and the link-atom approach

Systems considered in section 2.5.1 assumed non-bonding interactions between the QM and MM regions in a system, however this is rarely the case, particularly in enzymes and the condensed-phase systems that constitute the main body of this work. Therefore, QM/MM simulations must be able to treat atoms covalently bonded through a method interface. It is simply not enough to cut off the bonds and leave them unsatisfied in terms of electronic and orbital considerations. Instead, a number of approaches have been formulated to treat the problem, and will be briefly discussed here.

It is particularly important that although there may be bonds between both systems, the entire structure should behave as if a partition had not been imposed. Deviation from supposed "normal" behavior of a system due to partitioning reduces the reasonableness of the approach. Overall, there are three groups which encompass the solutions to the partitioning problem for covalently-bonded systems.

Firstly, with methods involving localized orbitals, hybrid orbitals can be formed for the QM atoms of the QM-MM bonds that have been broken through partitioning. These hybrid orbitals are parameterized in the direction of the cutoff MM atom, with the effective MM orbital constituting of a single electron, but satisfying the bonding requirement of the lone QM atom. However, it is important that the presence of the partition bond be incorporated into the MM forcefield by an extra bonding term. [166-172]

An alternative to the hybrid orbital method involves capping atoms. These replace the atom from the broken bond with a small moiety such as a hydrogen, or halogen. This atom is placed in the position of the MM atom from the broken bond, where it is therefore

not necessary to include an MM bonding term in the force field. The fact that the QM method does not need to be edited to incorporate the non-standard hybrid orbitals is certainly an advantage for this method, however the behavior of the system with the capped atoms must not deviate from the original, therefore the cap atoms must be chosen carefully to reproduce the properties of the partitioned bonds.

Finally, dummy or link atoms are employed, to replace the covalent bonds between the QM and MM atoms. Hydrogens are usually employed as link atoms, this differing from the capping methodology insofar as the link atoms are able to move, and are not assumed to have the same position as their MM counterparts. The link atoms are positioned along the partitioned bond with the bond being treated by an MM term. This methodology has been applied widely in QM/MM techniques as is the least difficult to incorporate into systems and largely gives reasonable results in terms of property replication.

Overall, it must be remembered that in order for the boundary and partitioned bonds to be reliably represented, each technique requires that the QM and MM influences be balanced to reflect the properties of the full system in each case.

2.5.4 Combining QM/MM with Molecular Dynamics

In the previous sections, the principle of considering large systems was introduced in terms of the QM/MM methodology. These large systems such as enzymes and protein environments, however, necessarily have a significant range of motions available, leading to different conformations which can affect reactivity. In order to sample these, a simulation must involve some time evolution component, enabling the system to follow trajectories based on simple forces.

Molecular dynamics (MD) solves this by employing the classical equations of motion for atoms and molecules in order to obtain the time evolution of a system. It is generally applied to many-particle systems and used classical mechanics to provide trajectories. These allow time-dependent phenomena such as diffusion, to be followed. In the case of the systems studied in this work, the COMT enzyme active site can consist of a number of SAM conformers, discoverable by MD. A straight QM/MM simulation would not necessarily obtain the average conformational structure and could lead to inaccuracies

in data obtained. This is why the ensemble averaging discussed further in Chapter 7 samples the QM/MM structures generated as snapshots of a timed MD simulation.

2.6 Performing Isotope Effect Calculations

One of the aims of the work carried out was to bring together a series of recommendations for future calculation of isotope effects, in order to provide some sort of continuity to results and to the field at large.

Currently, there are a number of programs and protocols available for calculation of isotope effects. These mostly depend on vibrational data and Hessians, but some work purely on the basis of thermodynamic data.[173] Applying modern concepts to the original formulation by Bigeleisen has recently been shown to greatly enhance the methodology, bringing it into the supramolecular age.

2.7 Gaussian Calculations and the Use of the SULISO Suite

Written by Prof. I. H. Williams in Cambridge and subsequently Bath and Castellon, the suite consists of four programs for vibrational characterisation and isotope effect calculation.

A coordinate and Hessian file; a *fort.7* output from Gaussian[174] is used as input for program camvib.[175] Gaussian calculations are carried out in different manners based on reactant/product structure searches, or transition structure searches. Gaussian can be obtained through licensing directly from Gaussian Inc., or through the following website:

http://gaussian.com/orders/

Gaussian optimises structures to minima based on the direction of the optimisation. Reactant and product structures for reactions considered within the scope of this thesis, are minima in all directions on the potential energy surface, therefore Gaussian includes functionality to calculate vibrational frequencies, also. These calculations allow the user to inspect the vibrational spectrum of the computed system for presence of imaginary

frequencies. Arising from the fact that TS are saddle points on the PES, the presence of imaginary frequencies indicates the order to which the saddle point is raised. This is equal to the number of directions of negative curvature at the stationary point corresponding to the TS. A first order saddle point indicates a TS in one direction only, and most frequently is the transient species of interest. The relative energy of the species is proportional to the number of imaginary vibrational modes it possesses; a second order saddle point would therefore be less energetically stable than a first order TS. Reactant and product structures are locate through traditional energy minimisation and geometry optimisation methods. TS are identified firstly through optimisation with the Berny algorithm.[142] The Berny algorithm functions by effectively following the largest negative contribution to curvature, whilst minimising the energy gradient. TS obtained through this method were subjected to frequency analysis to ensure that a single negative frequency was present, corresponding to the motion along the reaction coordinate. Calculations involving frequency analyses produced Hessian matrices in the form of Gaussian *fort.7* files, as discussed previously.

Inclusion of additional terms and calculation modules within Gaussian are called through specific keywords. The corrections employed to consider dispersion are called through *empiricaldispersion=GD3BJ*, however the nature of how these calculations are employed is the subject of a later chapter.

An example keyword line for a TS optimisation to a constrained cage structure is included below:

```
#P freq b3lyp/aug-cc-pvdz opt=ts punch=(coord,derivatives)
```

The first term, `#P`, requests that the program print additional output to the log file produced upon completion of the calculation. This is helpful in obtaining additional information regarding solvation, frequency, and optimisation parameters.

The `freq` keyword requests an analytical frequency calculation to take place, producing normal modes corresponding to the degrees of freedom of the molecule. The `b3lyp/aug-cc-pvdz` specification describes the electronic structure method to be used in the calculation. In this instance, the B3LYP functional, with the double-zeta augmented correlation consistent basis set is chosen. In order to request that the structure is optimised to a saddle point, the keyword `opt=ts` is included. As discussed above, this

employs the Berny algorithm to search for a saddle point along the principle direction of negative curvature. Finally, in order to print the *fort.7* file including the coordinates and Hessian for input to the SULISO suite, `punch=(coord,derivatives)` is specified. This produces as output additional to the traditional log file, a standard text file containing the coordinates, gradient of the energy, and Hessian matrix.

All electronic structure calculations were run on the BALENA high-performance computing facility at the University of Bath, using a total of 16 processors on a single node. Isotope effect calculations and post processing of data were run on the University of Bath LINUX facility, LCPU.

Written in FORTRAN77, camvib functions based on a series of keywords, referring to specific subroutines linked to different aspects of the vibrational calculation. Functionality includes force constant matrix printouts, spectra prediction and preparation for further treatment of the data in terms of isotope effect calculations.

The input file for camvib consists of the output *fort.7* from the electronic structure code, as well the keywords relevant to the computation. Incorporated within the input, is a valence coordinate specification, preceded by the Cartesian coordinates for the system, and followed by the Hessian matrix.

Output from camvib consists of a *fort.12* file containing coordinates and Hessian formatted for input to lipfr or cutoff. camvib uses the valence coordinate information in the input to construct a set of local symmetry coordinates, which are the basis for the vibrational frequency assignment to specific degrees of freedom. The spurious translations and rotations from an electronic structure code can be separated from the vibrational motion. The origin of the coordinate system must first be shifted to the centre of mass, and principal moments and rotations determined. By formulating coordinate vectors for rotation and translation around the principal moment of inertia, a projection matrix can be set up in order to effectively remove the set of vectors representing the translations and rotations (6 of these become exactly equal to zero in the Hessian). These can then be projected out in order to retain the non-zero eigenvectors in the Hessian. The output Hessian is therefore effectively “cleaned up”, and once diagonalised and mass-weighted, is converted into units of cm^{-1} to print out normal modes.[176]

For full, non-cutoff systems, the output camvib *fort.12* file is transformed into lipfr input. lipfr computes IPFRs for a full molecular system, corresponding to a stationary point on a potential energy surface (PES). lipfr uses the coordinates and Hessian, calculating the vibrational frequencies by mass-weighting the Hessian matrix according to the specified isotopologues and applying Equation 20, IPFRs are output. The isotope effect is then a simple quotient of heavy isotopologue over light; results can be easily combined in a spreadsheet.

2.8 Atomic subsets: applying the cutoff rule, and program UJISO

The benefit of considering subsets of molecular systems - in particular large supramolecular structures - was first discussed in Section 1.2. Practically, this involves the use of programs cutoff and ujiso, respectively.

cutoff is a small program written in FORTRAN, which provides the "cut-off" functionality for application to Hessian matrices output from second derivative calculations on large systems with electronic structure codes. cutoff transforms the full system Hessian into a smaller matrix corresponding to a user-defined subset of atoms.

Program cutoff was first written by IHW at the University of Bath, and functions through deletion of rows and columns from the full molecular Hessian corresponding to atoms not to be retained. The resultant subset Hessian comprises of elements only relating to atoms marked by the user as retained.

Output from cutoff is formatted similarly to the *fort.7* file from a Gaussian09 electronic structure calculation, and with small edits can be formatted for input to program ujiso.

ujiso was written by IHW whilst working at the Universitat Jaume I in Spain, and has an equivalent function to lipfr, only for subsets. IPFRs are calculated from the subset Hessian output from a cutoff run. The subset Hessian consists of $3N_s$ elements, the lower triangle of which does not produce 6 zero eigenvalues upon mass-weighting and

diagonalisation. By considering all $3N_s$ degrees of freedom as harmonic vibrations, ujiso produced IPFRs for subsets of the original full molecular system.

This is particularly relevant for the supramolecular systems discussed in Chapter 7, where application of programs cutoff and ujiso provide information on the reliability of the cut-off approximation, and allow for a detailed look at the effect of dispersion on the supramolecular enzymic systems studied.

2.9 Summary

This chapter has considered the first step in obtaining kinetic isotope effects: generating reasonable and reliable molecular structures, and their accompanying chemical data. The basic theory underlying the IE codes used was detailed in Section 1.1. An overview of methods involved in the characterization of supramolecular systems has been provided. Beginning with a brief analysis of the programs used to obtain the isotope effects discussed in subsequent chapters; the importance of additional techniques not directly used in this work, but currently relevant, was also highlighted. Experimental details such as optimization routines and commonly used Gaussian keywords have been included. The SULISO suite and associated utilities can be obtained free of charge from the following GitHub repository:

https://github.com/pbw20/SULISO

The following chapter builds upon the introductory part of this work by setting the scene of the environmental influence on the sign and magnitude of isotope effects.

3. An initial analysis of environmental effects on isotope effects: a review

The bulk of this work was carried out during MChem of PBW, and published as a Research Article in *J. Phys. Chem. B.* (see Ref. 1). Figures and associated content within this chapter originate from Ref 1 and associated Supplementary Information. The impact of this is particularly important in the timeline of this research, and makes an appropriate introduction to the environmental effect aspect of the work.

The COMT enzyme catalyses the transfer of a methyl cation from AdoMet to a catecholate in an S_N2-like process. To fully explain the chemistry of the transfer itself, it is important to understand the behaviour of the methyl cation, and what effect – if any – the solvent itself has on the observed isotope effects.

The methyl cation is reasonably facile to characterise in terms of computational techniques. Previous work by Williams and coworkers had discovered an unlikely trend in isotope effects depending on small changes in computational methodology. Indeed, use of a united atom model to describe the PCM solvent cavity, showed an increase in equilibrium isotope effect (EIE) with dielectric of the implicit solvent (from EIE=1.294 in vacuum to 1.332 in PCM water). Applying a universal force field (UFF) cavity model within the PCM resulted in a *decrease* in EIE with magnitude of PCM dielectric (EIE=1.294 in vacuum to EIE=1.235 in PCM water). It is clear that a single trend must be correct, however it was important to identify which one.

The previous study from the early 2000s, by Williams and coworkers found the surprising trend in equilibrium isotope effects from vacuum to dielectric. In order to verify that this was simply not an artefact of the calculations carried out in Gaussian03, Gaussian09 calculations were carried out which confirmed this unusual behaviour. As the effect of solvent was part of our remit for investigation into susceptibility of isotope effects, it was important to gauge this on a general level by considering the average dielectric field.

The *solvent effects on isotope effects* investigated within the context of this work consist of the relationship between the solvent and the isotope effect under consideration in terms of environment, not reaction. Here the isotopic substitution is in the solute, not

in the solvent, and the hypothesis questions the effect the solvent itself has on the magnitude and direction of the IE in the reaction.

In the context of solvent effects on isotope effects, equilibrium isotope effects were used to analyse the effect of transferring from one medium to another; in this case, vacuum to solvent (Equations 27 and 28).[177]

$$CH_3{}^+(1) \rightarrow CH_3{}^+(\varepsilon) \quad (27)$$

$$CD_3{}^+(1) \rightarrow CD_3{}^+(\varepsilon) \quad (28)$$

Where ε and 1 represent the values for the dielectric constant.

The overall EIE for the change of dielectric is therefore expressed as:

$$\frac{CH_3{}^+(1) \rightarrow CH_3{}^+(\varepsilon)}{CD_3{}^+(1) \rightarrow CD_3{}^+(\varepsilon)} \mathrm{a} \frac{K_H}{K_D} = \frac{\left[CH_3{}^+(\varepsilon)\right]\left[CD_3{}^+(1)\right]}{\left[CH_3{}^+(1)\right]\left[CD_3{}^+(\varepsilon)\right]} = MMI \times EXC \times ZPE \quad (29)$$

Where MMI, EXC and ZPE refer to the mass-moment-of-inertia, vibrational excitation, and complex vibrational zero-point energy factors, respectively.

The equilibrium constant can equally be considered as a partition function (PF) ratio *i.e.* the probability of finding products over reactants. These PFs can be split into the individual components for the degrees of freedom with the product of the electronic, vibrational, translational and rotational contributions, yielding the PF.[178] Each contribution depends on different molecular parameters. Masses determine the translational PF, the rotational PF depends on mass and geometry, with the vibrational PF being governed by geometry, masses and force constants. A full description of the derivation and application of EIEs and IPFRs can be found in Chapter 1.

3.1 Method Validation

Before beginning a study on the effect of solvent on isotope effects using implicit solvation methods, it was important to gauge the relative reliability of the vibrational data obtained from a number of different and appropriate electronic structure methods.

Three test molecules were used as a set representing the methyl cation; the methyl cation itself, chloromethane, and the trimethylsulfonium cation. Each of these molecules had been vibrationally characterized experimentally, allowing scale factors to be calculated from the vibrational data output from each electronic structure method. The calculated vibrational frequencies were compared to the experimental results for each electronic structure method; that with the value closest to unity being chosen as the most reliable method.

Based on the study of the M06 and B3LYP functionals, as well as the AM1 semiempirical method, with a range of basis sets used for each functional, the method providing the scaling factor closest to unity for the test set of molecules was found to be B3LYP/aug-cc-PVDZ.

3.2 Solvent Effects on methyl cation isotope effects

Using this electronic structure method in Gaussian09 to minimize the methyl cation in vacuum and within PCM dielectrics ranging from 1.5 through to 80, the solvent cavity was defined separately as UFF and UA0 in order to verify the original Gaussian03 treatment. For each value of dielectric, the C-H bond lengths and angles were re-optimised and IPFR parameters re-evaluated, with the default radii for the UA0 and UFF methods being used, the results shown in Figure 5.

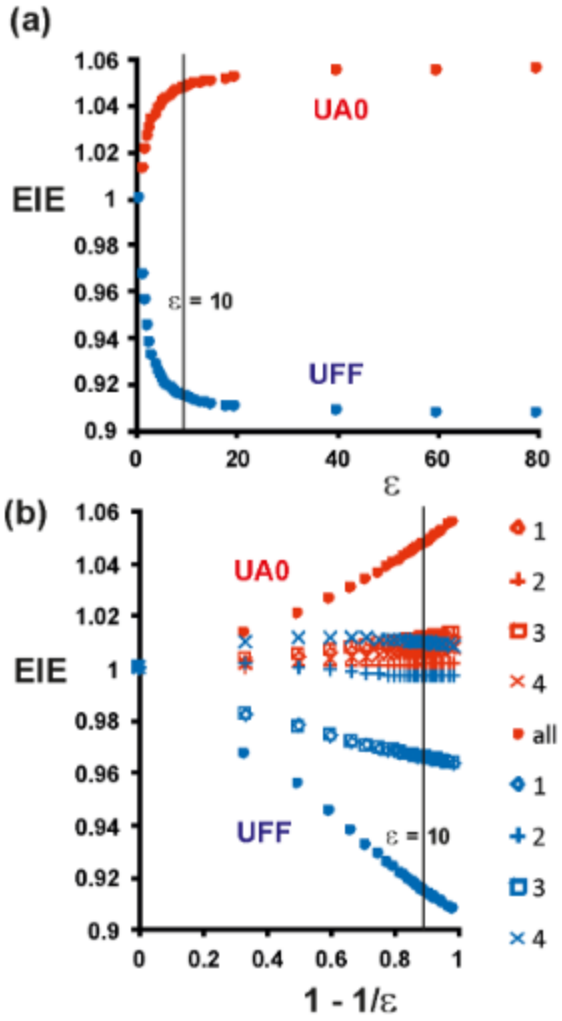


Figure 5. The UA0 and UFF radii-based EIEs for the methyl cation from vacuum to dielectric against (a) dielectric variation itself, ε, and (b) 1-1/ ε. Solid circles represent the full EIEs, ◊ the CH stretch, + the out of plane bend, □ the degenerate CH stretches, and × the degenerate in-plane bending modes. Extracted from Ref.1.

There is a definite trend towards quasi-opposite final values for the EIEs issued from UA0 and UFF cavity models, up to a dielectric corresponding to water (~80). Whereas the UA0 method predicts EIEs to follow an increasingly normal trend, finishing towards 1.06, the UFF method predicts EIEs becoming increasingly inverse with increasing solvent polarity, tending towards approximately 0.90. The issue remains that each method suggests an entirely different environment for the methyl cation based on the isotope effect: UA0 suggests the methyl cation is looser at increasing dielectric, UFF

suggesting a tighter nature in PCM solvent when compared to vacuum. The overriding finding however, is the dramatic dependence of the EIE on the dielectric constant, in particular in the region of 1 to 10.

It is worth highlighting that these trends are identical to those obtained in the earlier work with Gaussian03, and that differences between the previous Gaussian03 default (UA0) and current (UFF) have been noted on the basis of property relationships, by other workers.[179-181]

There is of course a charge component associated with this cationic system, for which the EIEs follow the trends shown in Figure 5(b). This corresponds to the $(1-1/\varepsilon)$ factor in the Born model for electrostatic contributions to the free energy of solvation of a point charge within a spherical cavity.[101, 182, 183]

It is unsurprising that the solvation energies relative to vacuum were found to be negative for both methods, their magnitude changing with respect to dielectric similarly to the EIEs. Indeed, it is the UFF solvation energy which is found to be larger by 56kj mol^{-1} than its UA0 equivalent at a dielectric of 80. It is noteworthy that again, the CH stretching modes (shown as ◊ in Figure 5) are the major contributors to the total EIE, to a sum of 114%, whereas the bending modes contribute only -14%.

The change which is exhibited is most clearly seen in the range of $1 \leq \varepsilon \leq 10$, with the majority of enzyme active sites having dielectric constants within this range. This has implications for enzyme-catalysed reactions in general; local values of dielectric in protein and enzyme environments can be manifest by charge redistribution, separation or neutralisation, resulting in KIE variations not solely due to the reacting system.

In order to gauge the reasonableness of the either cavity model, a more realistic description of the system was required. A QM/MM simulation of the methyl cation in a 40 Å box of water was carried out, with the AM1 semiempirical method describing the QM region, and TIP3P potential describing the water molecules. IPFRs were calculated for each of the snapshots taken from an MD simulation of the system, these going on to each be minimised by the QM/MM methodology detailed above.

Each of these effective poses constituted of a different arrangement of the solvent and solute, their average representing a conformationally, thermally averaged representation of the system. The CD_3 EIE for transfer of the methyl cation from vacuum to explicit

water gave a value of 0.85 ±0.02, *i.e.* an inverse effect. This can be rationalised by the protons/deuterons in the methyl cation experiencing a more condensed bonding environment in water than in vacuum, which is reflected by the valence force constants for each of the snapshot poses, showing increasing stiffness of the C-H(D) bonds on going from vacuum to dielectric.

It is worth noting that the EIE of 0.85 is quoted as an average of the locally-relaxed structures from the MD-QM/MM calculation consisting of 40 different solvent configurations. In terms of independent determination of the EIE in order to compare the UA0 and UFF results within the PCM, the QM calculations clearly support the prediction of the UFF method and are in complete disagreement with the trend exhibited by the UA0 cavity model. It is important to note the relative low computational expense of the AM1 model, which may not be entirely quantitatively definitive, however certainly supports the individual sphere model as opposed to a more unified approach.

3.3 UFF and UAO behaviour

An analysis of the valence force constants for the CH bonds in the methyl cation structures for each value of PCM dielectric showed analogous behavior to that of the isotope effect; increasing solvent polarity resulted in looser bonds for UA0 and stiffer bonds for UFF, corresponding to the normal and inverse EIEs, respectively.

When comparing these force constants with the bond length changes with solvent, the optimized CH bond lengths became shorter in the UFF model over a wide range, compared to lengthening in the UA0 model, but over a shorter range. The solvation energies for each structure were computed and plotted against bond length. UFF gave a linear correlation, the methyl cation exhibiting increased stabilization with decreased CH bond length. Conversely, the UA0 solvation energy varied increased quadratically with increased CH bond length, this being attributed to the UA0 cavity specification, which forms a large overall cavity centred on the carbon atom of the methyl cation, and encompassing each of the hydrogen atoms. By default, this provides less functionality of movement for the atoms, the CH bond being inherently constrained by the UA0 cavity where the H atoms approach the boundary at the limit of the cavity radius.

Conversely, the UFF cavity benefits from individual spheres centred on each atom, allowing for independent extension of the bonds from the central atom, unhindered by cavity specification.

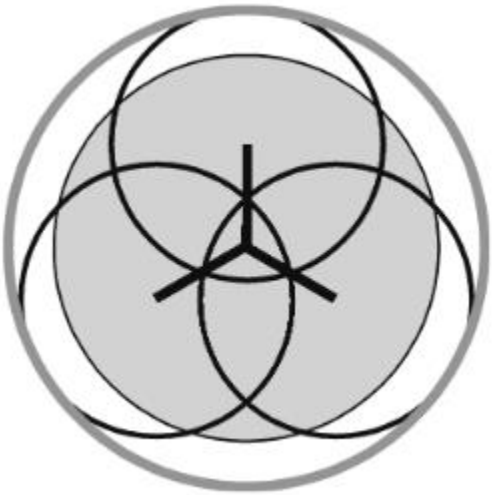


Figure 6. Model of the UFF and UA0 cavities as established from a methyl cation structure. The central divergent three lines represent the methyl cation, with the black empty circles the UFF radii of the hydrogen atoms, and the gray shaded circle the UFF radius of the carbon atom. The gray lined circle therefore represents the default UA0 radius for the methyl group. Extracted from Ref 1.

Due to the nature of the cavity models investigated, UA0 and UFF by nature result in different cavity shapes for the methyl cation in PCM solvent. Figure 6 shows the shape variations and therefore the different exposures to the solvent environment of the methyl cation through the medium of the solute cavity. The default UA0 radius (2.525Å) surrounds the carbon and each of the hydrogen atoms entirely as a sphere, large in size relative to the UFF sphere contributions. Considering that an average CH bond length is approximately 1.1Å, and that the UFF carbon and H radii are 2.04Å and 1.44Å, respectively, a radius increase in the UA0 model results in the EIE form vacuum to solvent transfer decreasing in magnitude and tending toward 1. This can be rationalised by the more vacuum-like effect of a large cavity surrounding the solute, while increasing the CH bond lengths within the UA0 cavity bring the hydrogen atoms closer to the cavity

boundary, and therefore the solvent. This has the effect of greater charge stabilization on the hydrogen atoms at the boundary limit, the H charge increasing proportionally with the decrease in carbon charge, leading to more negative solvation energies.

The EIE for transfer of the methyl cation from vacuum to solvent dielectric becomes less inverse then more normal for the carbon atom radius increase, while keeping the hydrogen radius at default. This can be considered as effectively turning the UFF cavity into a UA0 model by expanding the carbon to encompass the hydrogen cavities. This is true in the limit of a very large cavity, however the smaller hydrogen solute spheres still affect the interactions at radii of <3Å. Conversely, by varying the size of the hydrogen cavities in the UFF model, it was possible to gauge the effect the carbon atom plays on the magnitude and direction of the EIEs. Keeping the UFF carbon atom radius fixed, while changing the hydrogen atom radii, for values of hydrogen atom of less than 1.55Å, the UFF cavity has exposure to dielectric from both carbon and hydrogen atom spheres. However, for hydrogen cavity radii of more than 2.32Å, only the hydrogen atom cavities are exposed to dielectric, whereas in the range between 1.55 and 2.32 Å, both atom cavities are exposed to the dielectric in the plane of the methyl cation, but only hydrogen cavities exposed perpendicular. These structural features can be established from the calculated EIEs for each surface structure.

Beginning at the smaller values of hydrogen atom cavity radii, and increasing gradually, the EIE becomes inverse at first, before tending towards unity after reaching the 1.55Å limit discussed above. The noteworthy conclusion from this manipulation of cavity structures relates to the carbon atom effect on the total EIE. It appears that interaction of the carbon with dielectric, stabilizing the charge on the atom, is responsible for the inverse nature of the EIEs themselves. Conversely, decreasing the CH bond lengths and collapsing the hydrogen spheres onto the carbon resulted in an increase in carbon atom charge that was appropriately stabilized by its increased exposure to the dielectric effect.

When the CH bond length within the UA0 cavity causes the hydrogen atom charge to approach the cavity boundary, it experiences a greater degree of stabilization from the dielectric continuum. As the hydrogen atom charge increases, so does the stabilizing effect of the solvation energy, and a decrease in the CH stretching force constant. Furthermore, increasing the CH bond length in the UFF cavity results in an increased surface area of the hydrogen atom sphere, resulting in a decreased exposure of the

carbon atom to the dielectric continuum, leading to decreased stabilization and stronger bonds through larger CH stretch force constant. It was possible to confirm the UFF behavior with comparisons to other models, namely the SMD solvation model. [184] If the design of cavities remain spherical or CPK-like, the same electronic structure method with the PCM model replaced by SMD gave an EIE for transfer of the methyl cation from vacuum to PCM water as 0.90, in complete agreement with the UFF result.

3.4 Summary

EIEs calculated with the B3LYP/aug-cc-PVDZ method within the PCM and UFF cavity have been found to agree with AM1 QM/MM molecular dynamics simulations based on 40 structures of the methyl cation in TIP3P water. Indeed, the UFF cavity model predicts a trend of EIE becoming increasingly inverse (EIE<1) with increasing value of dielectric constant, tending towards approximately 0.90. This is in agreement with AM1/TIP3P QM/MM molecular dynamics simulations, which predict an EIE of 0.85±0.02 averaged over 40 solvent configurations for transfer of the methyl cation from vacuum to a solvent box of approximately 2000 water molecules. A trend opposite to that for the UFF model is observed when the UA0 cavity model is implemented in the simulations, leading to an increasingly normal EIE with dielectric, with a plateau at EIE~1.06. As the UA0 cavity model had been the default in versions of Gaussian earlier than Gaussian09, this has important consequences for the implementation of cavity models with PCM. In Gaussian09, UFF is the default cavity model. The difference in trends in EIE for each model can be attributed to the charge exposure of the solute in the cavity. Whereas UA0 places a solvent sphere only around the heavy atom(s), leading to a single sphere encompassing the methyl cation in this study, the UFF model places a sphere around each atom, leading to a more complex topology for the cavity. This allows the solvent to exhibit a stablising effect to a greater extent on the atomic charges within a UFF model as opposed to a UA0 cavity. A UFF model is therefore recommended for PCM calculations.

In this chapter, the factors affecting the isotope effects for solvation of the methyl group, and in essence, the effect of solvation on parameters inherently important in the calculation of isotope effects in general has been considered. It has been noted that the

EIE for solvation of the methyl cation in implicit solvent from a vacuum environment, changes significantly in the range of $1 \leq \varepsilon \leq 10$, which most enzyme active sites exhibit. The magnitude dependence of the isotope effect on solvation and environment is therefore important in order to correctly interpret results in a more general context. The influence of the environment must be considered in addition to purely structural factors, in order to obtain the correct interpretation of isotope effect data. The following chapter will further this concept by considering the method by which the frequencies are calculated for a molecule, and the most appropriate treatment for these.

4. Anharmonic effects on isotope effects and vibrational frequencies

Isotope effects being intrinsically linked to the vibrational properties of molecules, an investigation was carried out into factors influencing the direction and magnitude of calculated isotope effects by considering the vibrational approximations used. It is commonplace to incorporate the harmonic approximation in electronic structure theory, however in reality, vibrations of atoms in molecules are intrinsically anharmonic. Therefore, incorporating a study of electronic structure methods, as well as scaling factor determination, a publication co-authored by Prof. Ian H. Williams and PBW was produced in *Molecular Physics* on the effect of anharmonicity on isotope effect calculations, on which this chapter is heavily based. Additionally, consideration has been given to the effect of common combinations of parameterisations in electronic structure theory, on the vibrational data and calculated isotope effects, yielding recommendations not to combine implementations of PCM with anharmonic corrections in Gaussian09, especially with certain electronic structure methods (M06, B3LYP). The questions addressed by this study therefore revolved around the nature of anharmonic corrections, and their effect on the reliability/accuracy of isotope effect calculations. Additionally, the concepts of electronic structure method suitability and the effect of basis sets were considered for the test sets of molecules, treating the consequence of different sizes of basis function on the computed isotope effects.

As a development of the work on the PCM model and the effect of solvent on isotope effect, the effect of anharmonic corrections on isotope effects is a logical addition to our study of the theoretical basis of these subtle probes. By also incorporating hybrid PCM-anharmonic calculations, it was possible to offer strict guidelines on the calculation of IEs when using the tested methods. Indeed, the precise nature of anharmonic corrections on isotope effects has not been significantly treated in the recent literature, especially considering advances made in popular electronic structure software.

4.1 Background

Inductive isotope effects have been the cause of much research. Indeed, it was noted that an anomalous trend appeared in early computed EIEs for heterolysis of 2-propyl substrates with different leaving groups. These were interpreted by a model based on hyperconjugation (conformationally-dependant), and an inductive, conformationally-independent term.[185] These inductive effects can be better described by considering that anharmonic properties of bonds being studied. The effect of CH bond anharmonicity is the reduction in bond length in a carbon-deuterium bond, as compared to a carbon-hydrogen bond. This is due to the increase in reduced mass associated with larger isotopologous substitution causing the vibrational zero-point energy to be lower than that of its lighter equivalent in the anharmonic potential well.[24]

The vibrationally-averaged electron distribution is therefore different for these systems, leading to electrostatic interaction isotopic differences. Indeed, changes in harmonic force constants for vibrations can be considered as originating from minor perturbations such as substituent changes. These perturbations when acting upon the anharmonic potential result in a change in both the position of the energy minimum, as well as its localised curvature.

The Bigeleisen-Mayer theory of EIEs has had anharmonic corrections applied, when the equilibria under consideration had enough experimental vibrational data available. [15] However, Wolfsberg has pointed out that these anharmonic corrections offer little added value when compared to the harmonic approximation; K for deuterium interchange between water isotopologues was calculated to be 3.85 with anharmonic corrections, versus 3.84 under the harmonic approximation.[186] It is unsurprising that the default attitude for a number of years has been to employ harmonic partition functions in the calculation of EIEs and KIEs, given the relatively small pool of anharmonic data available.[187] Even with the continuous increase in computational capabilities, it is still rare to find anharmonic corrections being used frequently within the vibrational research and isotope effect community.[188]

Building upon the work carried out on methyl cation transfer from vacuum to dielectric, and the environmental effects observed on isotope effects, it was important to establish the significance of anharmonic corrections to this and the field in general. Granted, the

subject of anharmonic corrections as applied to methods with small basis sets has been treated in the scientific literature, with the majority of sources suggesting that it is unnecessary, especially when compared to scaled harmonic frequencies.[189] The previous study did not include anharmonic corrections for this reason, and the lack of experimental data on which to base an observed EIE.

Observed fundamentals for the methyl cation isotopologues were obtained, based on experimental work done in 2013 by Roueff and coworkers.[190] The group had characterised a number of modes for CH_3^+, CH_2D^+, CHD_2^+ and CD_3^+. In addition, recent application of vibrational configuration interaction (VCI) theory, provided meaningful anharmonic data on which to base a critical evaluation of the role of anharmonicity in partition functions for isotopologues of the methyl cation and for the EIE for methyl cation transfer from vacuum to PCM water.

As in Chapter 3, electronic structure calculations were carried out using Gaussian09 for both vacuum and solvent structures. The dielectric continuum describing the solvated structures was parametrised according to integral equation formalism PCM (IEFPCM), with the UFF cavity model being used for the solute and default atomic radii. The B3LYP and M06 hybrid density functionals, and second order Møller-Plesset perturbation theory (MP2) were used in conjunction with the 6-31+G(d), aug-cc-PVDZ, and aug-cc-PVQZ basis sets. These methods were chosen for their previous performance (B3LYP in our solvent effect study), general popularity and incorporation of parameterised dispersion (M06) and higher-order nature (MP2).

The above methods were each used for analytical second derivative calculations of harmonic frequencies. Slower than its numerical counterpart, analytical second derivative calculations are more computationally expensive but provide an exact answer, whereas the numerical equivalent often employed on earlier machines was much faster but more approximate by nature.

Anharmonic corrections were calculated using Barone's second-order perturbative method in Gaussian09. This treatment uses a finite-difference formula in the direction of the normal modes in order to obtain the third and a subset of the fourth derivatives.

Due to the use of anharmonic frequencies in the calculation of IPFRs, employing the Teller-Redlich product rule is unsatisfactory, due to its basis on the separability of

rotational and translational motion applying only within the harmonic approximation. Therefore the MMI factor can only be calculated based on Equation 16. The use of anharmonic frequencies in this work effectively represents an incarnation of simple perturbation theory, providing reasonable values for calculated IPFRs.[191, 192]

4.2 Results: B3LYP, MO6 and MP2 vibrational frequencies and scaling factors

Tables 2 through 5 include both harmonic and anharmonic frequencies, measured in wavenumber, for the isotopologues of the methyl cation from B3LYP, M06 and MP2 electronic structure calculations. Comparisons with the published VCI frequencies by Alcaraz mentioned in section 4.1 are shown.[193] For consistency, each mode is referred to within the C_{2v} point group for CH_2D^+ and CHD_2^+, with the same notation applied to the D_{3h} isotopologues CH_3^+ and CD_3^+. There have only been 9 modes of the total 20 which have been experimentally characterised, but which all agree well with the VCI results, therefore in order to maintain a consistent comparison, the so-called *Alcaraz* frequencies are used.

The basic harmonic frequencies calculated for each of the isotopologues shows a general tendency for decreasing r.m.s. error with basis set size. MP2 and B3LYP anharmonic frequencies give smaller r.m.s. errors in the frequencies than the harmonic equivalents, however this is not the case for M06. Calculation of the scaling factors was carried out by taking the sum of the results over the "best" calculated. The finite difference step size for numerical differentiation was maintained at the default (0.025 Å) for B3LYP, MP2 and M06 (with an additional step size being tested for M06 in Table 4) this being the recommended value for calculations.[194]

RMS errors were greater for the harmonic calculations for each of the three basis sets, with 6-31+G(d) providing the anharmonic frequencies with the lowest error and scaling factor closest to unity. In the context of the anharmonic frequencies, the scaling factors were used only as indicators of the validity of the approximation for that electronic structure method, and not as an explicit conversion.

Isotopologue	basis set		6-31+G(d)		aug-cc-PVDZ		aug-cc-PVQZ	
	frequency/ cm^{-1}	Alcaraz	harmonic	anharmonic	harmonic	anharmonic	harmonic	anharmonic
CH_3	v 1(a1)	2940	3070	2956	3013	2926	3021	2947
	v 2(a1)	1404	1422	1395	1427	1408	1425	1419
	v 3(a1)	3108	3265	3118	3220	3071	3211	3073
	v 4(b1)	1394	1425	1388	1385	1338	1411	1373
	v 5(b2)	3108	3265	3114	3220	3065	3211	3066
	v 6(b2)	1394	1425	1383	1385	1346	1411	1370
CH_2D	v 1(a1)	3002	3149	3015	3097	2960	3098	2971
	v 2(a1)	2240	2337	2235	2299	2205	2299	2213
	v 3(a1)	1389	1419	1377	1379	1333	1406	1364
	v 4(b1)	1308	1324	1296	1329	1289	1326	1302
	v 5(b2)	3107	3265	3111	3219	3061	3211	3064
	v 6(b2)	1171	1193	1150	1159	1091	1182	1149
CHD_2	v 1(a1)	3055	3211	3061	3164	3008	3159	3011
	v 2(a1)	2168	2250	2176	2211	2135	2214	2147
	v 3(a1)	1036	1054	1028	1024	994	1044	1018
	v 4(b1)	1205	1218	1189	1223	1181	1220	1197
	v 5(b2)	2356	2439	2357	2404	2319	2398	2325
	v 6(b2)	1281	1304	1264	1268	1219	1293	1254
CD_3	v 1(a1)	2097	2172	2024	2132	2035	2137	2074
	v 2(a1)	1090	1103	1033	1107	1061	1104	1066
	v 3(a1)	2345	2438	2354	2404	2318	2398	2319
	v 4(b1)	1030	1046	1024	1018	984	1037	1023
	v 5(b2)	2345	2438	2351	2424	2314	2398	2312
	v 6(b2)	1030	1046	1019	1018	992	1037	1020
RMS error			88.7	21.5	59.5	43.1	56.2	26.2
sum(frequencies)		46603	48278	46418	47529	45653	47651	46077
scaling factor			0.965	1.004	0.981	1.021	0.978	1.011

Table 2. B3LYP calculated harmonic and anharmonic frequencies for the methyl cation isotopologues, RMS error from benchmark frequencies, and calculated scaling factors. Alcaraz refers to the "best" calculated (CCSD(T*)-F12a/cc-pVQZ-F12) frequencies (Ref. 19).

Both of the correlation consistent basis sets underestimated the anharmonic frequencies to a greater extent than 6-31+g(d), which if considered as Irikura suggests,[189] should only be quoted to two significant figures. It follows that the 6-31+G(d) anharmonic results therefore show greater proximity to the Alcaraz "best" calculated (CCSD(T*)-F12a/cc-pVQZ-F12).

Isotopologue	basis set		6-31+G(d)		aug-cc-PVDZ		aug-cc-PVQZ	
	frequency/ cm^{-1}	Alcaraz	harmonic	anharmonic	harmonic	anharmonic	harmonic	anharmonic
CH_3	ν 1(a1)	2940	2998	3008	2953	2955	2982	2663
	ν 2(a1)	1404	1379	1463	1389	1448	1377	1056
	ν 3(a1)	3108	3216	3254	3182	3200	3184	2835
	ν 4(b1)	1394	1374	1447	1338	1618	1351	862
	ν 5(b2)	3108	3216	3189	3182	3059	3184	2851
	ν 6(b2)	1394	1374	1523	1338	1381	1351	1058
CH_2D	ν 1(a1)	3002	3087	3115	3047	2967	3064	2725
	ν 2(a1)	2240	2293	2311	2263	2323	2274	2094
	ν 3(a1)	1389	1368	1510	1332	1577	1345	1018
	ν 4(b1)	1308	1284	1389	1293	1351	1282	1006
	ν 5(b2)	3107	3216	3246	3182	3089	3184	2838
	ν 6(b2)	1171	1150	1279	1120	1326	1131	804
CHD_2	ν 1(a1)	3055	3157	3242	3120	2913	3129	2766
	ν 2(a1)	2168	2201	2234	2170	2203	2187	1985
	ν 3(a1)	1036	1016	1120	1190	1268	999	586
	ν 4(b1)	1205	1182	1313	989	1073	1180	1006
	ν 5(b2)	2356	2401	2421	2375	2434	2377	2145
	ν 6(b2)	1281	1256	1474	1225	1643	1237	1080
CD_3	ν 1(a1)	2097	2121	2047	2089	2034	2109	1808
	ν 2(a1)	1090	1069	1058	1077	1102	1067	724
	ν 3(a1)	2345	2401	2484	2375	2427	2377	2125
	ν 4(b1)	1030	1009	1080	983	1278	993	539
	ν 5(b2)	2345	2401	2401	2375	2290	2376	2129
	ν 6(b2)	1030	1009	1175	983	1039	992	754
RMS error			55.8	106.9	69.6	135.1	45.0	312.4
sum(frequencies)		46603	47178	48783	46570	47998	46732	39457
scaling factor			0.988	0.955	1.001	0.971	0.997	1.181

Table 3. M06 (default step size) calculated harmonic and anharmonic frequencies for the methyl cation isotopologues, squared differences from benchmark frequencies, and scaling factors.

Isotopologue	basis set		6-31+G(d)		aug-cc-PVDZ		aug-cc-PVQZ	
	frequency/ cm^{-1}	Alcaraz	harmonic	anharmonic	harmonic	anharmonic	harmonic	anharmonic
CH_3	ν 1(a1)	2940	2976	2954	2921	4033	2956	4784
	ν 2(a1)	1359	1381	1487	1396	2544	1380	3195
	ν 3(a1)	3108	3193	3953	3147	5278	3159	5875
	ν 4(b1)	1370	1373	557	1333	3272	1352	2877
	ν 5(b2)	3108	3193	2343	3147	3468	3159	3814
	ν 6(b2)	1370	1373	-1040	1333	2116	1351	3032
CH_2D	ν 1(a1)	3005	3065	2672	3014	3934	3038	4318
	ν 2(a1)	2240	2276	2500	2239	3719	2256	4449
	ν 3(a1)	1389	1367	-840	1327	3072	1345	2890
	ν 4(b1)	1299	1286	1061	1300	2562	1285	2821
	ν 5(b2)	3106	3193	2668	3014	3934	3159	4318
	ν 6(b2)	1171	1149	-109	1115	2738	1132	3015
CHD_2	ν 1(a1)	3056	3134	1269	3086	2552	3103	2691
	ν 2(a1)	2168	2185	2360	2147	3544	2169	4328
	ν 3(a1)	1036	1015	-128	986	1978	1000	2397
	ν 4(b1)	1188	1183	296	1196	2605	1182	2265
	ν 5(b2)	2356	2385	2743	2350	4179	2389	4738
	ν 6(b2)	1281	1257	-1312	1220	4181	1237	3523
CD_3	ν 1(a1)	2097	3105	1890	2067	2908	2091	3757
	ν 2(a1)	1085	1071	1023	1082	2129	1070	2596
	ν 3(a1)	2345	2384	2951	2350	4191	2359	4897
	ν 4(b1)	1030	1009	171	979	2796	993	2386
	ν 5(b2)	2345	2384	1349	2349	2334	2358	2774
	ν 6(b2)	1030	1008	-1418	979	1723	992	2662
RMS error			209.9	1211.2	217.8	1392.4	208.0	1674.0
sum(frequencies)		46482	47945	29400	46077	75790	46515	84402
scaling factor				0.969	1.581	1.009	0.613	0.999

Table 4. M06 (0.03Å step size) calculated harmonic and anharmonic frequencies for the methyl cation isotopologues, squared differences from benchmark frequencies, and scaling factors.

Isotopologue	basis set		6-31+G(d)		aug-cc-PVDZ		aug-cc-PVQZ	
	frequency/ cm^{-1}	Alcaraz	harmonic	anharmonic	harmonic	anharmonic	harmonic	anharmonic
CH_3	ν 1(a1)	2940	2976	2954	2921	4032	2955	4784
	ν 2(a1)	1404	1381	1487	1396	2544	1380	3195
	ν 3(a1)	3108	3193	3953	3147	5278	3159	5875
	ν 4(b1)	1394	1373	557	1333	3272	1351	2877
	ν 5(b2)	3108	3193	2343	3147	3468	3158	3814
	ν 6(b2)	1394	1373	-1040	1333	2116	1352	3032
CH_2D	ν 1(a1)	3002	3065	2672	3014	3934	3038	4421
	ν 2(a1)	2240	2276	2499	2239	3719	2256	4449
	ν 3(a1)	1389	1367	-840	1327	3072	1345	2890
	ν 4(b1)	1308	1286	1061	1300	2562	1285	2821
	ν 5(b2)	3107	3193	2668	3147	3923	3159	4318
	ν 6(b2)	1171	1149	-109	1115	2738	1132	3014
CHD_2	ν 1(a1)	3055	3134	1269	3086	2552	3103	2691
	ν 2(a1)	2168	2185	2360	2147	3544	2169	4328
	ν 3(a1)	1036	1015	-128	986	1978	1000	2396
	ν 4(b1)	1205	1183	296	1196	2605	1182	2265
	ν 5(b2)	2356	2384	2743	2350	4179	2359	4738
	ν 6(b2)	1281	1257	-1311	1220	4181	1237	3523
CD_3	ν 1(a1)	2097	2105	1890	2066	2908	2031	3757
	ν 2(a1)	1090	1071	1022	1082	2129	1070	2596
	ν 3(a1)	2345	2384	2951	2349	4191	2359	4897
	ν 4(b1)	1030	1009	171	979	2796	993	2386
	ν 5(b2)	2345	2384	1349	2350	2334	2358	2774
	ν 6(b2)	1030	1008	-1418	979	1723	991	2661
RMS error			43.0	1214.2	37.5	1405.8	36.4	1715.6
sum(frequencies)		46603	46944	29399	46209	75778	46422	84502
scaling factor			0.993	1.585	1.009	0.615	1.004	0.552

Table 5. M06 (0.001Å step size) calculated harmonic and anharmonic frequencies for the methyl cation isotopologues, RMS error from benchmark frequencies, and scaling factors.

Following the original B3LYP calculations, the same methodology was used in conjunction with the M06 functional (Tables 3, 4 and 5). The default step size (Table 3) was again used in the numerical differentiation procedure in order to obtain the anharmonic frequencies. Due to the relatively large magnitude of the r.m.s. errors, further calculations were carried out in order to investigate the effect of the step size on accuracy of the anharmonic corrections. The errors highlight a distinct difference

between harmonic and anharmonic scaling factors, the latter showing a greater r.m.s. error than the pure harmonic frequencies.

A change in the numerical differentiation step size: to 0.001 Å (Table 5) and 0.030 Å (Table 4) results in a small variation in the accuracy of harmonic frequencies compared to the benchmark, however anharmonic frequencies still show large r.m.s. errors. The 0.001 Å step size (Table 5) reduces the RMS error of the harmonic frequencies by half, but for the anharmonic frequencies; results in an factor of 10 increase in those for 6-31+G(d) and aug-cc-pVDZ, and a factor of 5 increase in those for aug-cc-pVQZ. Implementing larger step sizes (0.01, 0.03) results in both harmonic and anharmonic RMS errors increasing from the values for the default step, leading to a degradation of both sets of frequencies. Due to the consistent, relative accuracy of the harmonic values obtained from each basis set using the M06 functional, it is unlikely that the functional itself is responsible for the erroneous results when implementing the anharmonic correction, but instead that the *combination* of the functional with the anharmonic approximation is producing artefacts in the computations which result in the unusual anharmonic frequencies obtained. It has been discussed previously that the vibrational perturbation method (VPT) of Barone and coworkers can result in some discrepancies in degenerate frequencies when combined with certain functionals.[195-197] This is particularly evident with the frequencies obtained with M06.

In relation to the degenerate mode splittings; whereas B3LYP showed a variation of 3-8 cm^{-1} between the ν_3/ν_5 stretching and ν_4/ν_6 in-plane bending modes, the magnitude of the M06 anharmonic frequencies did vary significantly, up to approximately 200cm^{-1}. Within at least Revision A.02 of Gaussian09, the implementation of anharmonic corrections for the M06 hybrid meta exchange-correlation functional beaves anomalously. The functional itself performs as well as other methods for the harmonic frequencies, and energy scans for the structures displaced along the normal coordinate for each component of the degenerate CH_3^+ stretch fit well with fourth-order polynomials with R^2 values of more than 0.999. The force constants from these scans equate to 3.405 and 3.42o aJ $Å^{-2}$, but correspond to a frequency split themselves of 65cm^{-1} for M06 compared to 4cm^{-1} for B3LYP.

Isotopologue	basis set		6-31+G(d)		aug-cc-PVDZ		aug-cc-PVQZ	
	frequency/ cm^{-1}	Alcaraz	harmonic	anharmonic	harmonic	anharmonic	harmonic	anharmonic
CH_3	v 1(a1)	2940	3143	3035	3086	3005	3090	2991
	v 2(a1)	1359	1462	1446	1452	1444	1445	1427
	v 3(a1)	3108	3356	3210	3305	3162	3298	3162
	v 4(b1)	1370	1472	1440	1436	1401	1451	1415
	v 5(b2)	3108	3356	3210	3305	3062	3298	3162
	v 6(b2)	1370	1472	1440	1436	1401	1451	1415
CH_2D	v 1(a1)	3005	3229	3102	3175	3049	3174	3055
	v 2(a1)	2240	2398	2311	2357	2261	2357	2276
	v 3(a1)	1389	1465	1433	1430	1395	1444	1409
	v 4(b1)	1299	1361	1347	1352	1329	1345	1329
	v 5(b2)	3106	3356	3206	3305	3158	3298	3159
	v 6(b2)	1171	1232	1208	1202	1165	1214	1188
CHD_2	v 1(a1)	3056	3298	3156	3245	3106	3241	3109
	v 2(a1)	2168	2306	2236	2265	2197	2266	2202
	v 3(a1)	1036	1088	1069	1062	1041	1073	1051
	v 4(b1)	1188	1253	1240	1244	1225	1238	1224
	v 5(b2)	2356	2507	2429	2468	2390	2463	2395
	v 6(b2)	1281	1347	1320	1314	1285	1328	1298
CD_3	v 1(a1)	2097	2224	2091	2183	2094	2186	2066
	v 2(a1)	1085	1133	1086	1126	1091	1120	1081
	v 3(a1)	2345	2506	2424	2467	2387	2462	2386
	v 4(b1)	1030	1081	1073	1055	1035	1066	1056
	v 5(b2)	2345	2506	2424	2467	2387	2462	2386
	v 6(b2)	1030	1081	1073	1055	1035	1066	1056
RMS error			149.1	70.3	113.0	37.6	112.1	39.9
sum(frequencies)		46482	49632	48009	48792	47105	48836	47298
scaling factor			0.937	0.968	0.953	0.987	0.952	0.983

Table 6. MP2 (default step size) calculated harmonic and anharmonic frequencies for the methyl cation isotopologues, squared differences from benchmark frequencies, and scaling factors.

In order to sample multiple electronic structure methods, MP2 perturbation theory was used with the same basis sets as B3LYP and M06 to calculate the harmonic and anharmonic frequencies for each of the methyl cation isotopologues. As with B3LYP, including the anharmonic corrections results in a factor of 2 – or more- improvement in the RMS error for the calculated frequencies, and hence a scaling factor closer to unity. Overall, the MP2 method overestimates the frequencies, both harmonic and

anharmonic, with the anharmonic values remaining further from the benchmark than their B3LYP equivalents.

Using the MP2 method with anharmonic corrections results in a preservation of the degeneracy of the $\upsilon 3/\upsilon 5$ stretching and $\upsilon 4/\upsilon 6$ in-plane bending modes. In contrast, the DFT methods B3LYP and M06 both cause these to split, B3LYP by a minor range of 5 cm^{-1}, M06 a more significant and variable range.

In Chapter 3, the vibrational zero-point energy dominance on the D_3-EIE for transfer of the methyl cation from vacuum to PCM water was discussed. These ZPE contributions are related to sums of vibrational frequencies, as described in Chapter 1. It is useful to employ the ratio of $(\Sigma\nu)_{VCI}/(\Sigma\nu)_{calc}$ to provide an alternative metric for the error in frequencies calculated by our methods, while also providing a uniform scaling factor which modifies the calculated harmonic frequencies towards the best available anharmonic values. It is common for more approximate electronic structure methods to overestimate vibrational frequencies, resulting in scaling factors of <1.[198, 199] This is observed for the harmonic frequencies from all methods and basis sets included in this chapter, with the sum of the MP2 anharmonic frequencies also overestimating for all basis sets with respect to the VCI benchmark values. The M06/aug-cc-PVQZ anharmonic sum of frequencies however seems significantly underestimated, where the B3LYP functional performs relatively well across the basis set range. Anharmonic vibrational frequencies with the B3LYP functional and the 6-1+G(d) basis set give a $(\Sigma\nu)_{VCI}/(\Sigma\nu)_{calc}$ = 1.001, and an r.m.s. error of 21.5cm^{-1}, which together are the scaling factor closest to unity, and smallest deviation from VCI values, respectively. B3LYP/6-31+G(d) was therefore chosen as the method of choice to estimate the domination of EIEs by vibrational ZPE.

4.3 EIEs: anharmonic corrections to IPFRs

Table 7 shows the IPFRs calculated with the B3LYP/6-31+G(d) method. The IPFRs for the 6 combinations of the methyl cation hydrogen isotopologues from harmonic and anharmonic calculations, are included. VCI results refer to the data published by Alcaraz and coworkers with a vibrational configuration interaction methodology using

coupled cluster theory and an explicitly correlated basis set for the quoted species in vacuum. MMI factors for each IPFR do not depend on the equilibrium bond lengths in the trigonal planar cation, with the IPFRs themselves (when defined as *heavy* isotope / *light* isotope) being characterised by values larger than unity. The magnitude of these IPFRs is temperature-dependant, becoming larger as temperature decreases, with the values in Table 7 agreeing with previous QM/MM work in an AM1/TIP3P model. Isotopic substitution was carried out on the methyl group within *S*-adenosylmethionine in water, and similar values of IPFRs being produced.[200]

An IPFR is proportional to a Gibbs free energy, if taking the natural logarithm. In order to evaluate errors in the IPFR, it is therefore most reasonable to consider the error in the % *ln* IPFR. For the B3LYP/6-31+G(d) IPFRs in Table 7, the r.m.s. error for the IPFR (ln f_1) as compared to the VCI values, is 7.9% for the harmonic and 1.0% for the anharmonic calculations in vacuum. Application of a scaling factor of 0.963 from the method validation calculations however results in a harmonic IPFR r.m.s. of 0.56%. This has two main implications. Firstly, for relatively reliable IPFR estimation, it is unnecessary to employ computationally expensive methods to evaluate the vibrational frequencies anharmonically. Secondly, with an error of ~8% for the unscaled harmonic frequencies, the B3LYP method combined with the modest 6-31+G(d) basis set provides an already not unreasonable estimation of the vibrational frequencies.

It is timely to briefly discuss the incorporation of the rule of the geometric mean into our isotopologue calculations. The application of the rule to isotope effects states that there can effectively be no isotope effects *on isotope effects*. This means that the contributions to the free energy from isotopic substitution at two different sites is additive. This was postulated based on the approximate local character of isotope effects, for which the approximation holds. Practically, the rule states that the IPFR for two isotopic substitutions at two sites, is simply the product of the individual IPFRs for both sites.[201]

		Alcaraz VCI	harmonic unscaled		anharmonic		harmonic scaled	
isotopic substitution	MMI	f_1	f_1	% error ln(f)	f_1	% error ln(f)	f_1	% error ln(f)
$CH3^+ \rightarrow CD3^+$	0.268951	1026.01	1736.08	7.6	1430.29	0.03	1259.13	0.02
$CH_2D^+ \rightarrow CD_3^+$	0.432650	107.404	158.989	8.4	136.528	0.22	128.118	0.16
$CH_3^+ \rightarrow CHD_2^+$	0.401475	91.6421	131.074	7.9	98.6267	0.08	106.001	0.16
$CHD_2^+ \rightarrow CD_3^+$	0.669908	11.1959	13.2450	7.0	14.5020	2.3	11.8786	0.53
$CH_2D^+ \rightarrow CHD_2^+$	0.645835	9.59322	12.0036	9.9	9.41440	-0.20	10.7856	1.2
$CH_3^+ \rightarrow CH_2D^+$	0.621636	9.55279	10.9195	5.9	10.4761	0.97	9.8279	0.30
r.m.s. % error in ln(f)				7.8		1.0		0.56

Table 7. IPFRs for the methyl cation hydrogen isotopologues in vacuum at 298.15 K (B3LYP/6-31+G(d)). Table extracted from Ref 4.

Indeed, the isotopologues described by the IPFRs in Table 7 and the scaled and unscaled harmonic frequencies from which they originate, all obey the rule of the geometric mean (Equation 30). Conversely, the anharmonic calculations do not (Equation 31).

$$f_1(CD_3/CH_3) = 1736.08 \approx 12.0036^3 = (f_1(CHD_2/CH_2D))^3 \quad (30)$$

$$f_1(CD_3/CH_3) = 1430.29 \neq 9.4144^3 = (f_1(CHD_2/CH_2D))^3 \quad (31)$$

Table 8 shows the IPFRs at 289.15K for the methyl cation in PCM water within the UFF cavity model. The IPFRs are larger than those in vacuum, which can be attributed to the CH stretching frequencies in solvent as discussed in Chapter 3. The ratio of IPFRs in vacuum to solvent, the EIE, f_1/f_{80} has no experimental equivalent, and would be inherently hard to characterise, in particular due to the highly reactive nature of the methyl cation, particularly in a water solvent.

The unscaled EIE for transfer of the methyl cation from vacuum to PCM dielectric was calculated as 0.908, compared to 0.911 for scaled harmonic. These values agree with the previous calculations discussed in Chapter 3, in particular with the SMD solvation model. Indeed, a similar conclusion can be drawn from these results, with the CH stretching frequency being affected by the stiffer bond in solvent, described by the larger stretching force constant.

It is reasonable to consider that EIEs for a pair of isotopologues which vary in mass number by two or less, will be proportionally closer to 1, than their more mass-variant equivalents. The rule of the geometric mean is briefly discussed in the context of cumulative isotope effects. It is also sometimes quoted as “isotope effects are independent and cumulative”,[187] however observation of the EIEs shown in Table 8 present a disagreement with this statement. Indeed, the EIEs of $K(CH_3)/K(CH_2D)$ and $K(CH_2D)/K(CHD_2)$ show remarkable similarity in magnitude, however $K(CH_2D)/K(CD_3)$, and $K(CH_3)/K(CHD_2)$ and $K(CH_2D)/K(CD_3)$ vary quite significantly.

	harmonic unscaled		anharmonic		harmonic scaled	
isotopic substitution	f_{80}	EIE	f_{80}	EIE	f_{80}	EIE
$CH3^+ \rightarrow CD3^+$	1911.00	0.908	1343.97	1.064	1381.89	0.911
$CH_2D^+ \rightarrow CD_3^+$	168.698	0.942	252.804	0.540	135.651	0.944
$CH_3^+ \rightarrow CHD_2^+$	141.628	0.925	48.4810	2.034	114.216	0.928
$CHD_2^+ \rightarrow CD_3^+$	13.5002	0.981	27.7216	0.523	12.0989	0.982
$CH_2D^+ \rightarrow CHD_2^+$	12.4960	0.961	9.11938	1.032	11.2118	0.962
$CH_3^+ \rightarrow CH_2D^+$	11.3338	0.963	5.31625	1.971	10.1871	0.965

Table 8. EIEs for transfer of the methyl cation from vacuum to PCM water (B3LYP/6-31+G(d)). Table extracted from Ref. 4.

Inspection of the EIEs originating from the anharmonic frequency calculations shows a considerable range both of magnitude and directions. Appearing between 0.5 and 2.0, these values of EIE suggest a combinatorial problem, much like that with M06, of the anharmonic corrections applied within Gaussian09, with the PCM model. The vacuum anharmonic and scaled harmonic values of $\ln f_1$ show excellent linear correlation with the equivalent Alcaraz VCI values. $\ln f_1$ for the anharmonic calculation corresponded to a slope of 1.04 and r^2 value of 0.997, and scaled harmonic, a slope of 1.03 and r^2 value of 0.9997. Additionally, the PCM scaled harmonic $\ln f_{80}$ when plotted against the Alcaraz VCI results yields a slope of 1.09 and r^2 value of 0.9997.

When comparing these good agreements with advanced methodological results, to the values for $\ln f_{80}$ obtained within the PCM incorporating anharmonic corrections, there is a distinctly reduced degree of correlation. The plot of anharmonic $\ln f_{80}$ in PCM against VCI $\ln f_1$ gives a slope of 1.09, but r^2 of 0.90, indicating the lack of correlation within the results, as well as the scatter. Furthermore, it has been pointed out that PCM in Gaussian09 indeed produces a cavity surface which is both continuous and therefore allows for differentiation in the direction of the normal modes in order to obtain anharmonic frequencies, whilst the choice of cavity model and size are inherently important.[202]

Indeed, the impact of cavity size and shape in the calculation of vibrational frequencies and isotope effects has been discussed at length in Chapter 3. From this, the decision to use the UFF cavity was taken, and produces reasonable results for harmonic frequency calculations. Of course, variations of cavity size and shape are intrinsically possible within Gaussian09, however it is clear that based on the current default values, the conclusions of the current study with regards to EIE calculations from anharmonic corrections as applied to the methyl cation in PCM water must highlight the preference for scaled harmonics, both in relation to the reliability of results, as well as computational expense.

4.4 Energy scans of vibrational modes of the methyl cation

Displacements along the normal coordinates for each of the vibrational modes of the methyl cation yield energy surfaces corresponding to the effect of the displacements on the molecular structure and its stability. Using the BATHMAT program written by PBW at Bath, the unit vector of each non-mass-weighted normal Cartesian coordinate was multiplied by successive iterations of 0.02 Å, and a sum carried out with the original coordinates from the equilibrium geometry. Over a range of ±0.4Å around the minimum of the potential energy surface, single point energies were calculated at each point, with regression to a fourth-order polynomial carried out, producing a very reasonable fit for each mode with R^2>0.999.

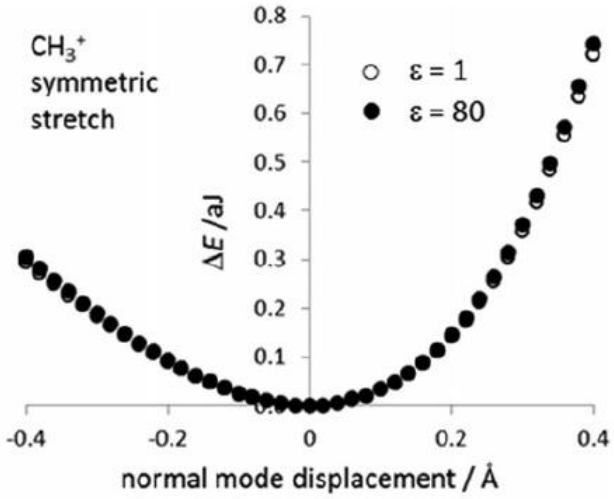


Figure 7. Plot showing single point energies calculated with the B3LYP/6-31+G(d) method for displacements in the direction of the symmetric stretch normal mode of the methyl cation in vacuum (open circles) and PCM water (full circles). Figure extracted from Ref. 4.

The quadratic, cubic and quartic coefficients of the respective regression equations are shown in Table 9. Figure 7 shows the anharmonic energy curves for displacement in both directions along the symmetric C-H stretch normal mode in both vacuum (open circles) and PCM water (full circles).

Normalised by a factor of √3, displacements in a negative direction indicate a bond length decrease, and those in a positive direction a bon length increase, over a total range of 0.46Å. Due to the negative sign of the cubic coefficient, the energy rises more significantly for compressions than extensions to the bond length. Magnitudes of the coefficients to the fourth order polynomial fit to the single point energies are almost identical in vacuum and PCM dielectric, however the harmonic force constant in PCM water, which is equal to twice the quadratic coefficient, is larger than in vacuum, these being 5.756 aJ $Å^{-2}$ and 5.576 aJ $Å^{-2}$, respectively. Indeed, this stiffening of the CH bond in solvent again agrees with the conclusions from Chapter 3, from a study of the UFF cavity within PCM dielectric.

Furthermore, the salient conclusion from section 4.4 is the inherently normal results obtained from the curve-fitting exercise for PCM solvation within Gaussian09. The PCM model describes the anharmonic potentials well for each of the normal modes, however this reaffirms the point that implementation of the corrections within Gaussian09 themselves, in combination with the PCM produce erroneous results, certainly in the context of the present study.

Table 9. Coefficients corresponding to fourth order polynomial fits to single point energies (B3LYP/6-31+G(d)) for displacements of Δx in the directions of the normal modes for the methyl cation in ε=1 (vacuum) and ε=80 (PCM water). Table extracted from Ref. 4.

	$\nu_1(a_1)$	$\nu_2(a_1)$	$\nu_3(a_1)$	$\nu_4(b_1)$	$\nu_5(b_2)$	$\nu_6(b_2)$
Coefficient of Δx^2 / aJ Å^{-2}						
$\varepsilon = 1$	2.788	0.659	3.463	0.738	3.479	0.659
$\varepsilon = 80$	2.878	0.656	3.603	0.741	3.600	0.655
Coefficient of Δx^3 / aJ Å^{-3}						
$\varepsilon = 1$	−3.348	−0.158	4.104	0.0	0.0	0.0
$\varepsilon = 80$	−3.440	−0.162	4.257	0.0	0.0	0.0
Coefficient of Δx^4 / aJ Å^{-4}						
$\varepsilon = 1$	2.388	0.051	6.567	0.123	6.315	0.045
$\varepsilon = 80$	2.473	0.055	6.687	0.128	6.482	0.055

4.5 Summary

In section 4.2, a publication by Jacobsen was surveyed, suggesting that anharmonic vibrational frequency calculations are not worthwhile for methods with small basis sets.[189] The work described in this chapter certainly agrees – in the most part – with the conclusions drawn by Jacobsen and coworkers. Indeed, it has been shown that the computational expense of anharmonic calculations is not necessarily justified for electronic structure methods using small basis sets, such as those included in this chapter.

The B3LYP/6-31+G(d) method is recommended for the calculation of partition functions and isotope effects, produces excellent anharmonic frequencies when compared to those obtained from the high-level VCI method of Alcaraz and coworkers. Moreover, scaled harmonic frequencies in vacuum produce further reliable IPFRs as compared to the more expensive anharmonically-corrected equivalents. The recommendation for condensed media calculations is again scaled harmonic frequencies, but from a PCM implementation within a dielectric continuum. Indeed, both harmonic and anharmonic vibrational frequencies were calculated for isotopologues of the methyl cation, both in vacuum, and PCM dielectric (where ε=80). B3LYP, MP2 and M06 electronic structure methods were using in combination with the 6-31+G(d), aug-cc-PVDZ, and aug-cc-PVTZ basis sets in order to calculate scaling factors based on comparisons with published 'gold standard' vibrational configuration interaction (VCI) results for CH_3^+, CH_2D^+, CHD_2^+, and CD_3^+. The scaling factor, calculated as the ratio of the sum of VCI frequencies to the sum of frequencies from individual combinations of electronic structure method and basis set, was determined and showed erratic results for M06 combined with anharmonic corrections for all basis sets. MP2 and B3LYP perform reasonably with all basis sets, however B3LYP/6-31+G(d) provides the scaling factor closest to unity for the methyl cation isotopologues. This method was used to consider the effect of implementing the PCM on IPFRs for methyl cation isotopologue transfer from vacuum to dielectric (ε=80). Harmonic scaled and unscaled PCM calculations provide a reasonable estimate of the EIE for transfer into solvent medium, however combined with anharmonic corrections, the PCM calculations give erratic results.

Isotope effect calculations based on harmonic frequency calculation produce reasonable and generally reliable results, and have been carried out since the dawn of computational chemistry. From the work presented in this chapter, it is recommended that it is not worthwhile to carry out anharmonic vibrational frequency calculations for isotope effects, as the employment of scaled harmonic frequencies as an alternative, offers advantages in terms of computational expense. Indeed, scaling factors can be considered as simple corrections for the implicit insufficiencies in electronic structure calculations such as electron correlation and truncated basis sets. Another factor covered by the scaling factor implementation, is the neglect of anharmonicity in most methodologies, insofar as if harmonic force constants describe changes in energy accurately for small displacements from equilibrium geometries, the ZPVE levels for two isotopologues where the same anharmonic potential has been applied to both cannot be harmonically reproduced to an equivalent level. Stating that scaled harmonic IPFRs improve the results for isotope effect calculations, implies that pure harmonic treatments for IPFRs are somewhat lacking and lead to some inaccuracies. However, it is important to remember that the harmonic approximation itself is reasonably applied in the calculation of isotope effects, with the caveat that the relevant force constants should represent the effect of anharmonicity themselves.

Concluding within this chapter: the data and arguments presented, suggest that anharmonic corrections are unnecessary for IPFR calculations from analytical second derivatives output from popular electronic structure programs, however anharmonicity itself should not be ignored, and can be incorporated by far computationally-cheaper means such as scaled harmonic frequencies, as compared to anharmonic corrections within electronic structure codes. The impact of this work on the general theme of this thesis: model system analysis to supplement our understanding of supramolecular interactions, is twofold. Firstly, an anharmonic analysis of supramolecular systems would be enormously computationally expensive, therefore it is appropriate that harmonic scaling factors can be used to approximate the effects of anharmonicity (as well as electron correlation) in the calculation of IPFRs for reactions. Additionally, the deviations to the rule of the geometric mean, presented in this chapter suggest that in fact isotope effects are not consistently cumulative, having important implications on the complex interactions and substitutions often present in supramolecular systems.

5. The effect of explicit solvation on Kinetic and Equilibrium Isotope effects

5.1 Introduction

The work in *J. Phys. Chem. B* was novel in its consideration of the solvent as having an effect in itself. By using a common average field implementation (PCM) suggestions were offered regarding the inherent effects solvents can impart on the solute. In order to develop our study and offer additional guidance particularly for interpretation of these effects in complex heterogeneous environments such as enzyme active sites, it was important to move to a more explicit definition of the solvent. This model would therefore be able to treat a series of questions which were as yet unanswered. As an inherently constrainable model, it was possible to specify H-bond distances explicitly, thereby observing the effect of small changes in the environment on the methyl cation and the isotope effect associated with methyl transfer. In addition, by varying the donor-acceptor distances in the direction of the reaction coordinate, the effect of a relatively compressed or expanded reaction environment could be considered. Moreover, this tests previous hypotheses relating to small changes in isotope effect being attributed to donor acceptor distances, as opposed to consideration of the remainder of the reaction environment (interactions perpendicular to the reaction coordinate *etc.*). Indeed, Klinman and colleagues reported T_3-KIEs on k_{cat}/K_m in hCOMT of 0.791±0.012 for the enzyme wild type, compared to 0.822±0.021 and 0.850±0.012 for the Y68F and Y68A mutants, which see residues in the donor-acceptor plane removed. This was considered as support for the compression hypothesis in terms of reaction coordinate compaction, however the degree to which alternative environmental factors affect these isotope effects, had not been explicitly considered. This study therefore treats both the donor-acceptor/reaction coordinate plane, as well as interactions perpendicular to this, reflecting effects in H-bonding and active site environment.

An explicit solvent environment, or *cage* as it will be referred to hereon in, was therefore designed. Although useful in terms of providing mean interactions, PCM and solvent models often fall short in describing those solvent-solute interactions on an individual scale. Therefore five water molecules were placed at the vertices of a methyl cation; three hydrogen bonding to the equator, and two acting as nucleophiles at the axes. From the

perspective of studies of the COMT enzyme, and claims made to support the compression hypothesis, not only was the solvent effect on isotope effect considered, but the effect of the proximity of the hydrogen bond donors to the methyl cation. This is of particular interest and resulted in the trend observed in a more constrained equatorial cage to an expanded one, effectively accounting for the compression effect quoted elsewhere.[92] The environment is the issue with common interpretations of both experimental and theoretical results in this situation. This means that it is often the donor-acceptor distances which are considered when attempting to rationalise isotope effect data, however, it has been demonstrated that factors perpendicular to the donor-acceptor axis can influence isotope effects similarly.

Throughout this chapter, the earlier work in Chapter 3 has been expanded upon, treating initial solvent effects on isotope effects and has been applied to enzyme-like test systems, providing additional suggestions for future isotope effect calculation and interpretation.[3] This is heavily based upon work published in *Angewandte Chemie,* coauthored by Prof. Ian H. Williams and PBW.

5.2 Isotope effects on the methyl cation within a constrained cage of water molecules

QM calculations for methyl transfer within a model "cage" system consisting of five water molecules situated at the vertices of a trigonal bipyramid, show that the 2^{o} D_3- and T_3-KIEs are inherently sensitive to hydrogen bonding interactions perpendicular to the plane of the methyl transfer axis.

The variations in hydrogen bonding distance (C-H···O interactions) were explicitly controlled, independently to changing the donor acceptor distances, in order to provide both a cumulative and combinatorial view of the solvent effects on KIEs within the cage structure. The application of this discovery to Adomet dependent methyltransferases supplements the work carried out by Trievel *et al.*[203], where the group concluded that CH···O hydrogen bonding in methyltransferases tends to be stronger than more traditional equivalents, with this interaction representing an important feature of AdoMet-dependent methyltransferases as a class of enzymes. This provides new evidence of the environmental influence on 2^{o} KIEs, suggesting that contributions other than compression in the donor-acceptor axis maintained by other groups, may be

responsible for the effects interpreted as supporting the compression hypothesis by others.[204]

As described in Chapters 1 and 3, methyl transfer itself occurs in numerous chemical and biological processes, from metabolism in the virus responsible for Dengue fever, to photochemical reactions.[205, 206] The question of methyl transfer TS stabilization relative to the RS by enzymes has caused much controversy (see Chapter 1, Section .5.4). A large inverse D_3 KIE for COMT-catalysed methyl transfer between AdoMet and catecholate, observed by Schowen and coworkers suggested to possibility of a compressive effect stabilising the TS.[207] Through QM/MM simulations which reproduced the same KIE as the original work, it was shown that there was no change in the mean donor- acceptor distance in the TS.[80, 90]

More recent experimental KIEs on mutant and wild-type COMT have been interpreted as renewed support for the compression hypothesis.[91, 208] Furthermore, it has been shown that hydrogen bonding influences in Adomet-dependent methyltransferases have important functional implications, their effect on KIEs remaining untested.[91] This chapter therefore concentrates on these unconventional effects in the environment surrounding the methyl cation, through a computational investigation of the isotopic sensitivity of the methyl cation placed within a constrained cage of 5 water molecules (Figure 8).

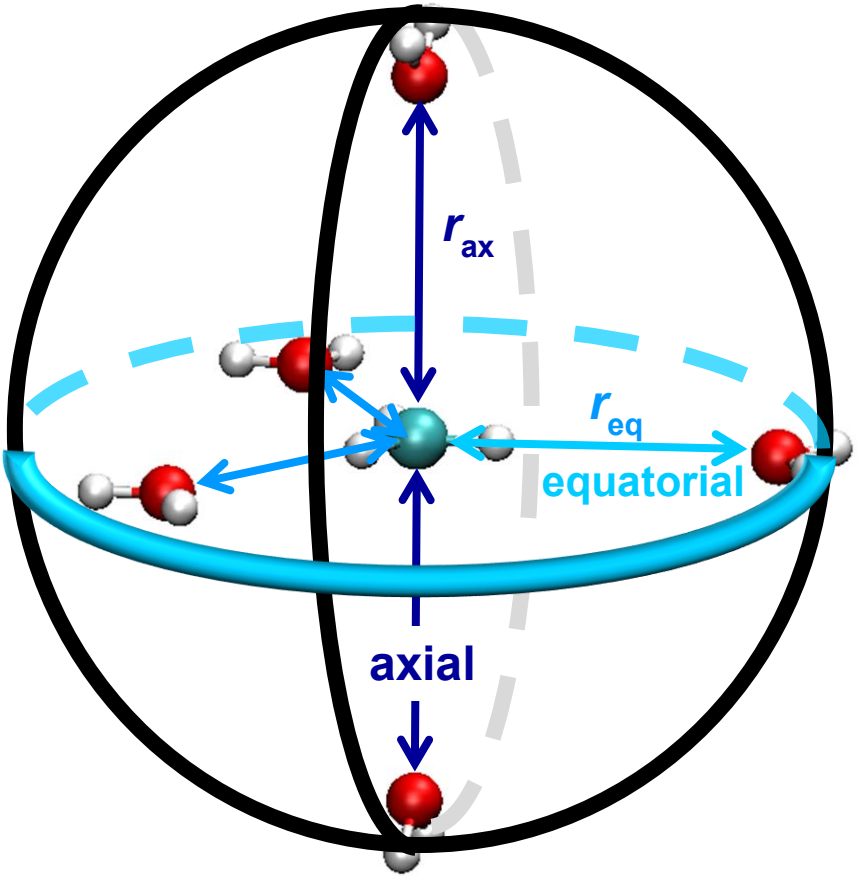


Figure 8. Structure of the constrained cage used within the explicit donor-acceptor and perpendicular interaction investigations. In this context, *axial* refers to the donor-acceptor axis, while *equatorial* denotes the CH···O, hydrogen bond distances. Figure extracted from Ref. 3.

The *axial* and *equatorial* distances shown in Figure 8 could be manipulated in a controlled manner, allowing the consideration of the individual axial/equatorial effects, as well as their combined influences on the KIEs produced. The importance of this model arises from its ability to investigate structures that would be experimentally complex to devise, thereby allowing us to derive significant structural data, which provides a framework for interpretation of these or similar interactions. Heterogeneous and constrained environments are not only unique to enzyme active sites, which are the motivation of the work presented in this chapter, but also nanoporous materials,[209-213] and sustainable chemical catalysts.[214-216] The work herein therefore offers meaningful

guidance on KIE interpretation within such constrained environments, highlighting the important effects for which non-reaction coordinate interactions are responsible.

5.3 Results and interpretation

The cage in Figure 8 consists of five water molecules at the vertices of a trigonal bipyramid. The structures with a methyl cation placed at the centre of the cage show *pseudo* D_{3h} symmetry, where the methyl group is coplanar with the three so-called *equatorial* water molecules, and perpendicular to the methyl transfer axis, described as the *axial* interactions hereon in. Coplanar with the equatorial water molecules, the methyl group therefore hydrogen bonds to each of these, which, like the axial water molecules, are frozen in space at user-specified values of r(C-H···O), or r_{eq}, and r(C···O), or r_{ax}, respectively. These are maintained in all RS and TS calculations, with the RS optimisations allowing the methyl cation to adopt a protonated methoxymethyl structure with an axial water molecule from the cage.

The B3LYP density functional was used, in accordance with the conclusions from Chapter 4, with the aug-cc-PVDZ basis set, which also performed well in the method validation calculations (Section 4.2), but more importantly is the same electronic structure method employed as the calculations in implicit solvent (PCM) in Chapter 3.

Isotope effects were calculated in the manner described in Chapter 1, with the KIE components and factor from the three C-H stretching frequencies determined by the Bigeleisen equation (Equation 17). A value of 0.85 for the deuterium isotope effect for transfer of the methyl cation from vacuum to solvent has been determined by a hybrid AM1/TIP3P MD simulation, as an average over 40 snapshots taken along the MD trajectory, these being locally relaxed within the QM/MM potential at 298K.[1]

The water molecules in the QM/MM simulations forming the first solvation shell around the methyl cation ranged in C···O distances from 2.95 to 3.20 Å, which roughly equates to the sum of the van der Waals radii. Moreover, transfer of the methyl cation in vacuum to the centre of the solvent cage, which is the methyl transfer TS by definition, produces a deuterium isotope effect of 0.86 for a symmetric cage structure where $r_{ax} = r_{eq}$ = 3.0 Å at 298 K (Table 10). This confirms the reasonableness of the cage model in terms of

replicating – at least in part – the interactions in more complex and realistic procedures. A decrease (compression) in r_{ax} from 3.0 to 2.0 Å increases the inverse magnitude of the isotope effect from 0.86 to 0.30, with r_{eq} remaining at 3.0 Å. This can be attributed to the loss of external degrees of freedom being compensated by gains in vibrational motions in the more constrained environment of the r_{ax} = 2.0 Å cage structure.

r_{ax} (Å)	r_{eq} (Å)				
	3	3.5	3.75	4	5
2.04	0.301	0.338	0.347	0.351	0.356
2.525	0.636	0.661	0.674	0.680	0.685
3	0.863	0.858	0.865	0.869	0.872

Table 10. EIEs for the transfer of the methyl cation from vacuum to the centre of the 5-water cage, for varied distances of r_{ax} and r_{eq}.

In order to understand the cumulative interactions within the cage structure, it was important to be able to model the explicit effects of the axial and equatorial substituents individually. Therefore, symmetric structures were minimized, the axial structure corresponding to Figure 9, where the axial distance can be varied and the equatorial substituent distance can effectively be deemed to be equal to ∞.

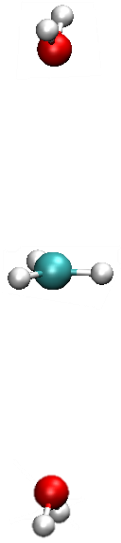


Figure 9. Axial structure corresponding to r_{ax} = 3.0 Å and r_{eq} = ∞.

One imaginary frequency for the transfer of the methyl group between axial water molecules is present in the minimized axial cage structures. Figure 10 shows the deuterium EIEs for transfer of the axial structures from vacuum into the centre of the complete solvent cage of Figure 8 *i.e.* effectively establishing the effect of incorporating the equatorial substituents from a distance of ∞ to 4Å and less. Figure 10 represents these EIEs as a function of r_{eq} for different values of r_{ax}. It can be seen that a change from 4 Å to 3 Å in r_{eq} for r_{ax} = 3.0 Å does not affect the EIE significantly. However the same change for r_{ax} = 2.0 Å, which is very similar in r(C···O) to the gas-phase S_N2 transition structure, increases the magnitude of the inverse isotope effect from 0.99 to 0.84; a significant change in EIE for a change in equatorial interactions.

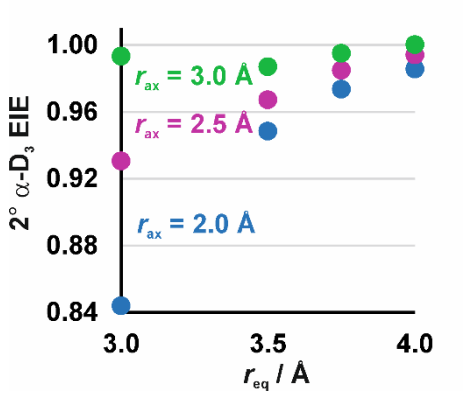


Figure 10. 2° α-D_3 EIES (298 K) for transfer of axial [H_2O…CH_3^+…OH_2] structures from vacuum into the center of the equatorial ring of the constrained cage. Figure extracted from Ref. 3.

r_{eq} / Å	r_{ax} / Å	$f_{complex}$	EIE	MMI	EXC	ZPE
3.0	2.04	69189	0.844	1.320	0.909	0.702
	2.525	32891	0.931	1.319	0.976	0.723
	3.0	24213	0.993	1.319	1.039	0.724
3.5	2.04	61558	0.948	1.322	0.888	0.806
	2.525	31643	0.967	1.321	0.931	0.786
	3.0	24366	0.987	1.321	0.980	0.762
3.75	2.04	59969	0.974	1.323	0.866	0.848
	2.525	31074	0.985	1.321	0.900	0.828
	3.0	24172	0.995	1.321	0.941	0.800
4.0	2.04	59233	0.986	1.323	0.846	0.878
	2.525	30790	0.994	1.322	0.873	0.861
	3.0	24046	1.000	1.322	0.904	0.837
∞	2.04	58381				
	2.525	30608				
	3.0	24052				

Table 11. B3LYP/aug-cc-pVDZ isotopic partition function ratios, 2° α-D_3 isotope effects for insertion of symmetric axial structures [H_2O…CH_3^+…OH_2] from vacuum to the center of the constrained three-water equatorial cage at 298 K and isotope effect factors.

Because each symmetric axial structure [H_2O…CH_3^+…OH_2] is a TS both with and without the equatorial waters, each isotope effect for transfer into the cage is an EIE. However, the IPFR ($f_{complex}$) for each structure includes a quantum correction on the transition frequency for methyl transfer which (approximately) cancels in the ratio EIE = $f(\infty)$ / $f(r_{eq})$. Again, the normal MMI factor is almost constant but for each value of r_{ax} the EXC becomes less inverse as r_{eq} decreases and the ZPE factor becomes more inverse and dominates the trend in the EIE. Some of these data are presented graphically in Figure 10. Transition structures were optimised using the Berny algorithm within Gaussian09 on each of the 25 constrained cage structures. Reactant/product structures were similarly optimised, but to a minimum. 2° α-D KIEs for the transfer of the methyl cation from one axial protonated methanol to the other are included in Figure 11.

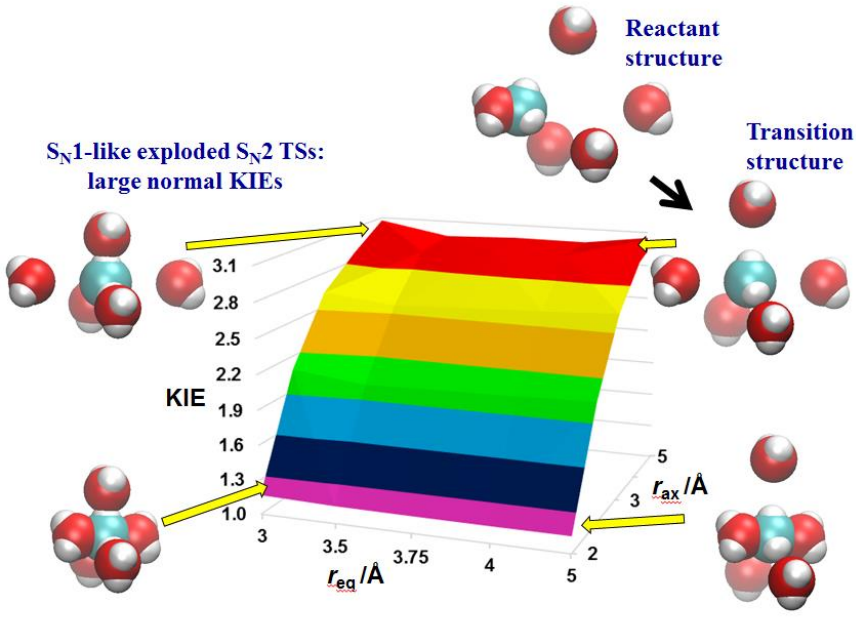


Figure 11. Changes of KIE with axial and equatorial distances of water molecules for the transfer of the methyl cation between each axial substituent. r_{ax} and r_{eq} refer to the axial C···O and equatorial C-H···O distances, respectively. Diagrams show TS geometries corresponding to each of the four extremities in the range of KIEs observed for changes in *r*. Figure extracted from the following contributed talk: I. H. Williams and P. B. Wilson, Proceedings of the 23rd IUPAC Conference on Physical Organic Chemistry, Sydney, Australia, 2016. *"Isotope Effect Calculations in the Supramolecular Age"*.

Figure 11 shows the variation of the KIEs with axial and equatorial distances of the solvent cage. Although there is slight variation as equatorial distance is increased, the vast changes in KIEs are observed from the changes in axial distance. This is unsurprising as the vector along the axial plane contains the reaction coordinate for the methyl transfer, therefore changes in geometry will have far more dramatic effects. Upon going from the most condensed structures to the exploded S_N2 structures, a gradual and significant increase in the magnitude of the normal KIEs is observed.

Large, normal KIEs are expected for the least hindered geometries, which reflect an almost S_N1-like character in the TS, such as in the r_{axial} = 5 results. Here, the methyl cation is almost a free carbocation with hydrogen-bonding interactions with the equatorial water molecules, and behaves as such. Conversely, the smallest normal effect is observed in the most condensed, r_{axial}=2.04; $r_{equatorial}$=3.0, structure. These findings are particularly relevant within the current climate of *compression* discussions, where it is shown that both axial and equatorial interactions have an effect on the KIEs, however these are heavily dominated by the axial distances.[3]

For each value of r_{eq}, a change in donor-acceptor distance results in the D_3 KIE from RS to TS within the cage, changing from approximately 1.1 at r_{ax} = 2.0 Å to approximately 3 at r_{ax} = 4.0 Å. This can be explained by the change from a tight S_N2 transition structure, to a looser, more *exploded* S_N2 TS.[217, 218] Reducing these KIEs to the magnitudes per deuterium, it is found that for r_{ax} = 2.0Å, and r_{ax} = 4 Å, the values are 1.03 and 1.4, respectively, which are reasonable magnitudes for 2° α-D_3 KIEs in the literature.[12]

The S_N1-like behaviour observed for the exploded cage structures can be attributed to the constraints imposed within the cage environment, which is relatively unseen for methyl transfers, and would be intrinsically difficult to study experimentally; it has however been noted in earlier computational work.[219]

		r_{eq} / Å				
r_{ax} / Å		3.0	3.5	3.75	4.0	∞
2.04	E_{RS}	-421.86098	-421.86284	-421.85963	-421.85637	-192.49373
	E_{TS}	-421.85431	-421.85515	-421.85132	-421.84755	-192.48296
	$\Delta E^{\ddagger}$	17.5	20.2	21.8	23.2	28.3
2.525	E_{RS}	-421.85795	-421.85585	-421.85265	-421.84965	-192.48948
	E_{TS}	-421.83256	-421.82884	-421.82388	-421.81939	-192.45370
	$\Delta E^{\ddagger}$	66.6	70.9	75.5	79.4	93.9
3.0	E_{RS}	-421.85239	-421.84793	-421.84511	-421.84261	-192.48528
	E_{TS}	-421.81262	-421.80504	-421.79898	-421.79371	-192.42612
	$\Delta E^{\ddagger}$	104	113	121	128	155

Table 12. B3LYP/aug-cc-pVDZ optimised energies for RS and TS structures (total energies in hartree, relative energies in kJ mol^{-1}). r_{eq} represents the equatorial C-H···O distance in Angstroms, and r_{ax} the axial C···O distance, also in Angstroms. E_{RS} is the ground state energy for the reactant structures, and E_{TS} the energy of the transition structures. $\Delta E^{\ddagger}$ is the activation energy for methyl transfer for each combination of r_{ax} and r_{eq}.

Table 12 shows the RS and TS structure energies for the constrained cage systems. $\Delta E^{\ddagger}$ is the barrier height for methyl transfer between axial waters within the constrained cage, either with the three equatorial waters present (r_{eq} = 3.0, 3.5, 3.75 and 4.0 Å) or without them (r_{eq} = ∞). The value of this barrier increases steeply as the C-O distance (= 2 Å r_{ax}) increases; this may be rationalized simply in terms of the Principle of Least Nuclear Motion, as has been noted previously.[219] Conversely, the value of this barrier decreases as the CH…O distance (r_{eq}) decreases. This reflects a differential between the RS and TS energies for interaction between each of the symmetric axial structures [H_2O…CH_3^+…OH_2] and the three equatorial water molecules which serves to stabilize the TS more effectively for shorter, stronger CH…O hydrogen-bonding interactions. (These interactions give rise to catalysis within the constrained cage model.)

It is the stronger C-H···O hydrogen bonding interactions which cause the height of the energy barrier within the constrained cage to decrease from 23.2 kJmol^{-1} at r_{eq}=4.0Å to 17.5 kJmol^{-1} at r_{eq}=3.0Å (Table 12), which corresponds to an increase in rate by an entire order of magnitude. Wolfe and Weinberg who presented calculations of D_3 KIEs for intramolecular methyl transfer between oxygen atoms within a constrained template,[220] however their so-called *transverse compression* effect was repulsive, with no stabilising effects from hydrogen bond acceptors, the transverse force causing a geometric distortion of the methyl cation in the TS from collinearity along the axial coordinate.

Earlier work by Moliner and Williams identified D_3 KIEs for methyl transfer within a symmetric cyptand which exhibited C-H···O interactions, however due to the nature of the system, it was not possible to manipulate these distances independently of the reaction coordinate distances.[93]

Table 13 contains the wavenumbers for motions within the constrained cage structures. There are, by nature of the complex environment, additional imaginary frequencies to that expected for the first order saddle point represented by the TS. For the EIE for transfer of the methyl cation to the constrained cage in Figure 8, Table 13 shows little isotopic sensitivity in the spurious imaginary frequencies, suggesting a static contribution to the isotope effect from these.

r_{ax} / Å		r_{eq} / Å									
		3.0		3.5		3.75		4.0		∞	
	description	CH_3	CD_3	CH_3	CD_3	CH_3	CD_3	CH_3	CD_3	CH_3	CD_3
2.04	ax-wag	418*i*	417*i*	419*i*	418*i*	416*i*	415*i*	412*i*	411*i*	369*i*	368*i*
		496*i*	**496*i***	**496*i***	**495*i***	**495*i***	**495*i***	**495*i***	**494*i***	**479*i***	**478*i***
	ax-wag	403*i*	403*i*	408*i*	408*i*	406*i*	405*i*	403*i*	402*i*	363*i*	362*i*
		295*i*	**295*i***	**280*i***	**279*i***	**274*i***	**274*i***	**269*i***	**269*i***	**137*i***	**137*i***
	ax-twist	146*i*	146*i*	122*i*	122*i*	113*i*	113*i*	105*i*	105*i*	82*i*	82*i*
		393*i*	**393*i***	**422*i***	**422*i***	**431*i***	**431*i***	**438*i***	**438*i***	**384*i***	**364*i***
	ax-twist	107*i*	107*i*	92*i*	92*i*	88*i*	88*i*	83*i*	83*i*	69*i*	62*i*
		54*i*	**53*i***								
	ax-transln.	56*i*	55*i*	34*i*	34*i*	48*i*	47*i*				
		61*i*	**60*i***	**16*i***	**16*i***	**35*i***	**35*i***	**48*i***	**47*i***		
	eq-wag	38*i*	38*i*			27*i*	27*i*	42*i*	41*i*		
		39*i*	**39*i***			**26*i***	**26*i***	**40*i***	**39*i***		
	eq-scissor	36*i*	34*i*			18*i*	18*i*	35*i*	35*i*		
		34*i*	**34*i***			**20*i***	**20*i***	**34*i***	**34*i***		
	flexing	32*i*	32*i*								
		22*i*	**22*i***	**44*i***	**44*i***	**46*i***	**45*i***	**50*i***	**50*i***		

Table 13 *continued*

r_{ax} / Å		r_{eq} / Å									
		3.0		3.5		3.75		4.0		∞	
	description	CH_3	CD_3	CH_3	CD_3	CH_3	CD_3	CH_3	CD_3	CH_3	CD_3
	rotation										
								4i	***4i***		
	MI	21.44705286		21.48090880		21.49458282		21.50631983		16.25090291	
		21.45034410		**21.48382664**		**21.49737106**		**21.50892768**		**16.32865153**	
	VP	21.44705287		21.48090880		21.49458279		21.50631972		16.25090292	
		21.45034411		**21.48382704**		**21.49737148**		**21.50892671**		**16.32865153**	
2.525	ax-wag	348*i*	347*i*	384*i*	383*i*	394*i*	392*i*	399*i*	398*i*	390*i*	389*i*
		500*i*	**499*i***	**502*i***	**501*i***	**502*i***	**501*i***	**502*i***	**501*i***	**482*i***	**481*i***
	ax-wag	342*i*	341*i*	380*i*	380*i*	389*i*	388*i*	394*i*	393*i*	376*i*	374*i*
		303*i*	**303*i***	**298*i***	**298*i***	**296*i***	**296*i***	**294*i***	**294*i***	**191*i***	**191*i***
	ax-twist	123*i*	123*i*	101*i*	101*i*	93*i*	93*i*	85*i*	86*i*	60*i*	60*i*
		48*i*	**47*i***								

Table 13 *continued*

r_{ax} / Å		r_{eq} / Å									
		3.0		3.5		3.75		4.0		∞	
	description	CH_3	CD_3	CH_3	CD_3	CH_3	CD_3	CH_3	CD_3	CH_3	CD_3
	ax-twist	85*i*	85*i*	72*i*	72*i*	70*i*	69*i*	68*i*	67*i*	43*i*	38*i*
		416*i*	**416*i***	**442*i***	**442*i***	**451*i***	**451*i***	**457*i***	**457*i***	**402*i***	**381*i***
	eq-twist	**73*i***	**73*i***			**34*i***	**34*i***	**60*i***	**56*i***	**56*i***	**50*i***
	sym-stretch	**57*i***	**57*i***			**54*i***	**53*i***	**54*i***	**53*i***	**51*i***	**50*i***
	eq-stretch					52*i*	52*i*	59*i*	58*i*		
						48*i*	**47*i***	**51*i***	**50*i***		
	eq-stretch					49*i*	48*i*	54*i*	53*i*		
						46*i*	**45*i***	**45*i***	**44*i***		
	eq-stretch					29*i*	29*i*	40*i*	40*i*		
		22*i*	**22*i***	**27*i***	**27*i***	**28*i***	**30*i***	**39*i***	**39*i***		
	MI	21.45869210		21.48940694		21.50172858		21.51231609		16.27344552	
		21.43062013		**21.46510757**		**21.47885354**		**21.49083006**		**16.35852703**	
	VP	21.45869201		21.48940646		21.50172858		21.51231604		16.27344552	
		21.43062013		**21.46511162**		**21.47898918**		**21.49082882**		**16.35852703**	
3.0	wag	234*i*	233*i*	311*i*	310*i*	339*i*	336*i*	357*i*	356*i*	416*i*	415*i*

Table 13 *continued*

r_{ax} / Å	r_{eq} / Å									
	3.0		3.5		3.75		4.0		∞	
description	CH_3	CD_3	CH_3	CD_3	CH_3	CD_3	CH_3	CD_3	CH_3	CD_3
	500*i*	**499*i***	**501*i***	**501*i***	**502*i***	**501*i***	**502*i***	**501*i***	**480*i***	**497*i***
wag	227*i*	227*i*	309*i*	309*i*	335*i*	335*i*	354*i*	353*i*	411*i*	408*i*
	308*i*	**308*i***	**307*i***	**307*i***	**306*i***	**306*i***	**305*i***	**305*i***	**218*i***	**218*i***
twist	105*i*	105*i*	89*i*	89*i*	82*i*	82*i*	76*i*	76*i*	51*i*	51*i*
	65*i*	**65*i***								
twist	64*i*	63*i*	58*i*	58*i*	55*i*	55*i*	52*i*	52*i*	34*i*	31*i*
	430*i*	**429*i***	**451*i***	**451*i***	**458*i***	**458*i***	**464*i***	**464*i***	**408*i***	**386*i***
sym-stretch	**34*i***	**34*i***	**35*i***	**35*i***	**36*i***	**36*i***	**36*i***	**36*i***	**36*i***	**36*i***
eq-stretch			32*i*	32*i*	67*i*	66*i*	70*i*	68*i*		
			43*i*	**43*i***	**42*i***	**39*i***	**48*i***	**47*i***		
eq-stretch			29*i*	28*i*	66*i*	65*i*	69*i*	67*i*		
			46*i*	**46*i***	**52*i***	**51*i***	**47*i***	**47*i***		
eq-stretch			41*i*	41*i*	46*i*	46*i*				
			43*i*	**43*i***	**40*i***	**40*i***				
flexing	29*i*	29*i*	16*i*	16*i*	16*i*	16*i*	17*i*	17*i*		
			19*i*	**19*i***						
MI	21.46950873		21.49761555		21.50879727		21.51839000		16.27623938	
	21.38917652		**21.42520186**		**21.43988590**		**21.45327838**		**16.34825736**	
VP	21.46950873		21.49761553		21.50879728		21.51839001		16.27623935	
	21.38917704		**21.42520188**		**21.43988594**		**21.45327900**		**16.34825736**	

Table 13. Imaginary vibrational wavenumbers (cm^{-1}) for motions of constrained cage water molecules. (Regular font correspond to the TS and **bold to the RS**.). r_{eq} represents the equatorial C-H···O distance in Angstroms, and r_{ax} the axial C···O distance, also in Angstroms. Each of the descriptions of modes reflects the motions within the constrained cage. *Sym-stretch* represents the symmetric stretching mode, and *eq-stretch* the equatorial stretching mode. MI and VP are the moment of inertia and vibrational product contributions to the IPFR, respectively.

Considering that neither the RS nor the TS structures correspond to stationary points on the potential energy surface for the full unconstrained system, and also because of the imposed symmetry, there are a number of additional modes associated with imaginary vibrational frequencies, all of which are presented in Table 13. Four of these are evident in the $r_{eq} = \infty$ structures without the three equatorial waters, and involve in-phase and out-of-phase combinations of wagging and twisting of the axial water molecules. It is obvious that the protonated methanol fragment of the constrained RS structures is unstable with respect to displacement towards a pyramidalized sp^3 O atom instead of the trigonal-planar geometry imposed by the constraints; this instability is manifest in both the RS and TS cage structures. The constrained structures represent saddle points for inversion and torsion of the axial water molecules, but in practice these modes are almost insensitive to isotopic substitution in the methyl group and therefore do not affect the calculated KIEs.

In the cage structures with r_{eq} = 3.0, 3.5, 3.75 and 4.0 Å, some of the modes involving the axial waters are strongly coupled with motions of the equatorial waters and the brief descriptions given in Table 13 are rather approximate. Imaginary frequencies are associated with some displacements of the equatorial waters away from their constrained positions. When r_{ax} = 2.04 Å, the [H_2O…CH_3^+…OH_2] fragment has no room for lateral motion with respect to the equatorial waters, but for r_{ax} = 2.525 and 3.0 Å, combinations of CH…O hydrogen-bond stretching motions are associated with imaginary frequencies: these modes may also be considered as translations of [H_2O…CH_3^+…OH_2] within the equatorial plane. The important observation, however, is that in practice these modes are almost insensitive to isotopic substitution in the methyl group and therefore do not affect the calculated KIEs.

According to the Teller-Redlich Product Rule (see Chapter 1)[14] these terms should be equal in value. The extent to which this is found to be true in practice provides a stern test of the correctness and internal consistency of the vibrational frequency calculations, and especially of the method employed to project out contaminating components of translation and rotation from the vibrational degrees of freedom. Inspection of the results presented in Table 13 shows agreement in all cases to the fourth decimal place, with the majority being to the seventh. This excellent agreement

is achieved only if the vibrational products are evaluated by treating all the imaginary frequencies as if they were real. It also serves to confirm the view that the presence of imaginary frequencies associated with vibrational modes of the constrained water molecules does not have any material influence upon the values (or indeed the validity) of the calculated KIEs.

		r_{eq} / Å				
r_{ax} / Å		3.0	3.5	3.75	4.0	∞
2.04	$(CH)_{RS}$	5.763	5.620	5.633	5.654	5.730
	$(CH)_{TS}$	5.863	5.752	5.779	5.812	5.926
	$\Delta^{\ddagger}$	0.100	0.132	0.146	0.158	0.196
	$\Delta\Delta^{\ddagger}_{eq}$	-0.096	-0.064	-0.050	-0.038	
2.525	$(CH)_{RS}$	5.539	5.502	5.524	5.549	5.626
	$(CH)_{TS}$	5.543	5.540	5.590	5.636	5.779
	$\Delta^{\ddagger}$	0.004	0.038	0.066	0.087	0.153
	$\Delta\Delta^{\ddagger}_{eq}$	-0.149	-0.115	-0.087	-0.066	
3.0	$(CH)_{RS}$	5.467	5.496	5.518	5.540	5.600
	$(CH)_{TS}$	5.320	5.376	5.440	5.495	5.654
	$\Delta^{\ddagger}$	-0.147	-0.120	-0.078	-0.045	0.054
	$\Delta\Delta^{\ddagger}_{eq}$	-0.201	-0.174	-0.132	-0.099	

Table 14. B3LYP/aug-cc-pVDZ average CH-bond stretching force constants (aJ $Å^{-2}$) for RS and TS for the methyl transfer within the contrained cage.

The environment of three methyl CH bonds within the constrained cage does not have strict 3-fold symmetry along the axial direction and so the individual CH stretching force constants are not quite equal. The values for the force constants presented in Table 14 are averages of the individual values for each of the three CH stretching force constants. For a particular value of r_{ax}, the value of the CH stretching force constant F tends to decrease with decreasing r_{eq}, except for a slight increase at r_{eq} = 3.0 Å for the shorter r_{ax} distances. As r_{ax} increases, the value of F decreases for both RS and TS, but more rapidly for TS, for all values of r_{eq}. Consequently the difference $\Delta^{\ddagger}$ between RS and TS changes from a positive number for r_{ax} = 2.04 Å (stiffer bond in TS) to a negative number for r_{ax} = 2.525 and 3.0 Å (looser bond in TS). The double difference $\Delta\Delta^{\ddagger}$ (between $\Delta^{\ddagger}$ values for each r_{eq} distance and $r_{eq} = \infty$) becomes more negative, monotonically, as r_{eq} decreases.

These values offer further confirmation that the conclusions from Chapter 3 are valid, regarding force constant changes in solvent media. Stiffer CH bonds are generally observed in more constrained systems than the looser, more exploded structures.

The KIEs for transfer of the methyl cation from vacuum to the constrained cage system where r_{eq} = 3.0 Å and the r_{ax} is varied, are shown in Table 15. It is important to consider both the explicit effect of the equatorial substituents as in Tables 10 and 11, however the transfer of the methyl cation to solvent has been the subject of previous work detailed in Chapter 3. Isotope effects for this transfer provide important information on the explicit effect of solvent on isotope effects themselves, the original calculations of course being carried out within a continuum model and not explicit solvation.

$f_{vacuum}(Me^+)$	20823				
r_{ax} / Å	$f_{complex}$	KIE	MMI	EXC	ZPE
2.04	69189	0.301	3.536	0.727	0.117
2.525	32891	0.633	3.538	0.669	0.269
3.0	24213	0.860	3.540	0.607	0.401

Table 15. B3LYP/aug-cc-pVDZ isotopic partition function ratios, 2° α-D_3 isotope effects for transfer of methyl cation from vacuum to the center of the constrained water cage with r_{eq} = 3.0 Å at 298 K and isotope effect factors.

Because each complex with the methyl cation in the center of the cage is a TS, these isotope effects are KIEs which include the quantum correction to the isotopic sensitivity of the transition frequency for methyl transfer. The normal MMI factor is almost constant and the inverse EXC and ZPE factors show opposite trends, but the latter dominates and is primarily responsible for the trend for a larger (more inverse) isotope effect with decreasing r_{ax} distance. The inverse nature of this isotope effect is unsurprising given that the transfer from vacuum to solvent has been shown to lead to such values.[1] Indeed, the effect of transfer of the methyl cation to the particularly condensed cage at r_{ax} = 2.0 Å produces a large inverse KIE of 0.31, which is again mainly dominated by the ZPE factor.

5.4 The *Superheavy Cage* approach

By establishing that the KIEs obtained for the cage of water are chemically reasonable, and validating the appropriate nature of the model system employed, it was necessary to convert the cage structure to a more protein-like environment, becoming of an enzyme active site. Therefore, in order to consider the possible effect of C-H···O interactions on KIEs in an enzyme active site, a *superheavy* constrained cage was created, where each hydrogen atom was given a mass of 999 Da. Given that the COMT enzyme has a total mass of 30 kDa, the approximation provided a reasonable model for a protein environment, also removing the spurious vibrational couplings between the methyl cation and the small hydrogen atoms on the water molecules by effectively dampening their motions.

r_{ax} / Å		r_{eq} / Å									
		3.0		3.5		3.75		4.0		∞	
		CH_3	CD_3	CH_3	CD_3	CH_3	CD_3	CH_3	CD_3	CH_3	CD_3
2.04	S	258	258	260	260	259	259	260	260	263	262
	A	339*i*	337*i*	357*i*	355*i*	363*i*	361*i*	367*i*	366*i*	385*i*	380*i*
2.525	S	82*i*	*82i*	85*i*	*85i*	84*i*	*84i*	83*i*	*83i*	78*i*	78*i*
	A	258*i*	248*i*	268*i*	259*i*	273*i*	264*i*	276*i*	267*i*	285*i*	275*i*
3.0	S	92*i*	92*i*	98*i*	98*i*	99*i*	99*i*	99*i*	99*i*	97*i*	97*i*
	A	190*i*	181*i*	197*i*	188*i*	202*i*	192*i*	206*i*	196*i*	216*i*	205*i*
Superheavy cage (all H atoms of "water" molecules have mass = 999):											
2.04	S	22	22	22	22	22	22	22	22		
	A	264*i*	257*i*	279*i*	273*i*	286*i*	280*i*	291*i*	285*i*		

Table 16. Vibrational wavenumbers (cm^{-1}) for symmetric and *antisymmetric* (italics) axial C-O bond stretching in transition structures of methyl cation in constrained water cages.

Presented in Table 16 are the vibrational frequencies for C-O bond stretching within the cage structure of Figure 9. Included within the Table are the results for 2.04 Å in the superheavy cage, in order to provide comparisons with the effectively *light* structural equivalent.

The antisymmetric stretching mode is the reaction-coordinate mode (or transition vector) for methyl transfer within the cage. Its frequency is imaginary (as expected) in all cases. Replacement of hydrogen in CH_3 by deuterium in CD_3 leads to a decrease in magnitude of this frequency (as expected). An increase either in the value of r_{eq} for a particular value of r_{ax} or in the value of r_{ax} for a particular value of r_{eq} leads to an increase in the magnitude of this frequency, consistent with an increase in the barrier height $\Delta E^{\ddagger}$ for methyl transfer (see Table 12).

The symmetric stretching frequency is insensitive to isotopic substitution in the methyl group because atomic motions of these atoms do not contribute to this normal mode of vibration. This frequency has a real value for r_{ax} = 2.04 Å but is imaginary for r_{ax} = 2.525 and 3.0 Å. However, it is important to recognise that these symmetric structures do not correspond to stationary points on the potential energy surface for the full unconstrained system: in particular, the gradient of the energy with respect to displacement in the symmetric stretching coordinate is not zero.

In the “superheavy” cage, all frequencies are lowered and the isotopically-sensitive modes of the methyl group are decoupled from those of the cage environment, which is the expected effect of a massive protein environment also. Table 17 includes the IPFRs, their factors and the KIEs for the transfer of the methyl cation within the superheavy cage of water molecules, from RS to TS.

		RS r_{eq} / Å				TS r_{eq} / Å				KIE r_{eq} / Å			
		3.0	3.5	3.75	4.0	3.0	3.5	3.75	4.0	3.0	3.5	3.75	4.0
D_3	IPFR	274385	23668	22907	22359	23617	20761	20183	19944	1.160	1.140	1.135	1.121
	QC									1.003	1.003	1.003	1.003
	MMI	0.336	0.332	0.331	0.331	0.347	0.342	0.342	0.342	0.969	0.969	0.968	0.969
	EXC	1.466	1.502	1.545	1.566	1.385	1.426	1.461	1.496	1.058	1.053	1.057	1.047
	ZPE	55581	47517	44835	43132	49343	42669	40561	39166	1.126	1.114	1.105	1.101
T_3	IPFR	4534493	3728891	3574577	3459265	3526567	2975164	2869461	2827638	1.286	1.253	1.246	1.223
	QC									1.006	1.006	1.006	1.006
	MMI	0.352	0.350	0.349	0.358	0.357	0.355	0.355	0.355	0.985	0.985	0.984	1.008
	EXC	2.010	2.101	2.196	2.234	1.836	1.933	2.009	2.082	1.095	1.087	1.093	1.073
	ZPE	6411169	5077315	4660873	4341584	5409525	4362471	4051919	3852361	1.185	1.164	1.150	1.127
^{13}C	IPFR	1.258	1.256	1.257	1.258	1.187	1.182	1.181	1.180	1.060	1.064	1.065	1.066
	QC									1.005	1.006	1.006	1.006
	MMI	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	EXC	1.045	1.052	1.053	1.054	1.030	1.036	1.037	1.037	1.015	1.015	1.015	1.016
	ZPE	1.204	1.194	1.194	1.194	1.158	1.147	1.145	1.145	1.039	1.042	1.042	1.043
^{14}C	IPFR	1.545	1.540	1.542	1.540	1.383	1.371	1.369	1.368	1.117	1.123	1.126	1.126
	QC									1.009	1.010	1.011	1.011
	MMI	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 17 (Page 119)**.** B3LYP/aug-cc-pVDZ isotopic partition function ratios and factors for RS and TS, and KIEs, for methyl transfer within a constrained superheavy cage with r_{ax} = 2.04 Å.

There is significant variation of the deuterium and tritium KIEs with equatorial C-H···O distance, with a 0.5 Å decrease in r_{eq} leading to an increase in the isotope effect of 2 and 3% respectively. Simply considering the three CH bond stretching modes, these contribute to the inverse magnitude of the isotope effect, as F_{CH} increases from RS to TS.[56] The factors associated with the KIES themselves become less inverse as r_{eq} decreases due to the decrease in the change in the force constant to the TS, $\Delta F^{\ddagger}_{CH}$. This is accompanied by an increase in the strength of the C-H···O hydrogen bonding interaction, leading to the proportional weakening of the CH bonds.

The factor associated with the CH bond stretch is responsible for the variation in KIEs with r_{eq}, itself being dominated by the ZPE factor. Moreover it is the contribution to the KIE of the factors from each of the other vibrational modes, which results in the overall normal direction of the isotope effects, in particular the bending motions of the methyl group within the constrained cage. Unsurprisingly, the CH bonds are shorter and stiffer in the TS structures, the environmental influences resulting in force constant increases on the RS bonds.

5.5 Summary

Using a model geometry for a methyltransferase system, the results obtained certainly indicate that the strength of C-H···O interactions can affect the magnitudes of D_3 and T_3 KIEs in reactions involving methyl transfer. Although using anionic hydrogen bond acceptors instead of the water molecules in this chapter may enhance the effect observed, the conclusions remain both valid and of direct importance to the field of methyl transfer investigation.

Previous studies have suggested the importance of hydrogen bonding in Adomet-dependent methyltransferases, however the link to isotope effects had not yet been made to interactions perpendicular to the reaction coordinate axis.[80]

Romesberg and coworkers have however shown the link between CD bond stretching frequencies and the local electric field within proteins.[221] The effects of the compression hypothesis have been presented as contributing to T_3 KIEs on k_{cat}/K_M in human catechecol-*o*-methyltransferase of 0.791 ± 0.012 for the wild-type and 0.822 ± 0.012 and 0.850 ± 0.012 for the Y68F and Y68A mutants, respectively.[91, 208] These results were interpreted as donor-acceptor mediated active site compaction, which from the model described in this chapter would equate to changes in r_{ax}. This chapter has however presented the important effects of environmental influences perpendicular to the methyl transfer axis, namely variations in r_{eq}. These have been shown to result in KIE variations remarkably similar to those interpreted as supporting axial compression, leading to support for the idea that factors other than donor-acceptor distance mediation can affect KIE variations observed in methyltransferases, in particular for COMT and its mutants.

Chapter 3 highlighted the sensitivity of 2^o-D_3 isotope effects to local dielectric constant; indeed the model described in this chapter suggests that interactions perpendicular and outside the direct methyl transfer axis must be considered when calculating isotope effects (beginning at the QM/MM boundary definition), and also in their interpretation. Previous calculations by other groups had included these interactions, albeit across the boundary of a QM/MM protocol, whereby the elements relating to these interactions with residues such as Met40 and Asp141, which are proximate to the hydrogen atoms in the methyl group within COMT, have been omitted from the Hessian.

It has been shown that an electrostatic origin for COMT-catalysed methyl transfer is in effect, and can contribute to describing the variations in KIEs obtained from calculation and experiment. Schramm has shown the importance of KIEs as tools within an interdisciplinary protocol in order to determine TS structure and guide the design of potential drugs as transition structure inhibitors.[222] The steps for KIE-based drug design are straightforward. Indeed, first the intrinsic KIE for the chemical step is measured, after considering the experimental conditions under which this is possible. KIEs are then computed from many electronic model chemistries until KIE data matching experiment is obtained. The electrostatic, steric and topological properties, as well as molecular descriptors can be used to design new inhibitors and drugs based on the KIE information from the simulations.[222] A notable example of this strategy was applied to the enzyme human purine nucleoside phosphorylase, or PNP, which catalyses the cleavage of purine nucleosides. Experimentally, PNP was first expressed in order to be produced with full substitution of stable heavy isotope such as ^{2}H, ^{13}C, and ^{15}N. This results in the formation of an enzyme with a mass roughly 10% that of the naturally abundant equivalent. With the production of this "heavy" enzyme, small protein motions at the lower end of the wavenumber spectrum become decoupled from protein motions. This is a similar strategy to that used in the context of this thesis for the *Superheavy* approach. Schramm has noted[223] that the KIE as k_{cat}/K_M is unchanged between "heavy" and naturally abundant enzyme, therefore the nature of the transition state will be the same for both isotopologous species, with additional information provided by the uncoupled motions.

It is well-known that computational techniques are more important than ever in contemporary drug discovery, however the value of the data obtained from chemical calculations are intrinsically dependent on both the computational methodology implemented, as well as the reliable interpretation of the information provided. This is particularly relevant to the quality of data obtained for transition structures in the context of enzymes and their reactivity, whereby the method used for KIE calculations in proteins is of significant importance. The work presented in this chapter has clearly demonstrated the influence of hydrogen bonding interactions perpendicular to the methyl transfer axis, on 2^{o} KIEs, and is of direct relevance to the modelling of other methyltransferases, but also the plethora of other supramolecular systems for which KIE data has been obtained experimentally.

6. An Evaluation of the Cutoff Procedure

6.1 Introduction

The traditional theory of KIEs based on molecular partition functions and the Bigeleisen equation, as described in Chapter 1, assumes that each molecular system, either RS or TS, corresponds to a zero-gradient stationary point on a potential energy surface. Reliable optimized geometries and Hessians are readily available from modern quantum-chemical packages employing second-derivative methods, provided that convergence thresholds are set suitably tightly. However, since the turn of the millennium, it has become common to perform simulations for very large molecular systems (either molecules in solution or within enzyme active sites) but to compute explicit Hessians for only a subset of the total number of atoms in the system. For example, relaxation of a specified subset of atoms to a local minimum (or saddle point) may be performed within a frozen environment of the remaining atoms. In the context of QM/MM methods (as reviewed in Section 2.5), the mobile subset and the frozen environment may be the same as the QM and MM regions, although different selections may also be made. In these circumstances the N_s subset atoms do not in themselves constitute a stationary structure in which vibrational degrees of freedom are separable from translations and rotation. The constraining influence of the environment means that diagonalization of the mass-weighted $3N_s$ x $3N_s$ Hessian in Cartesian coordinates generally yields $3N_s$ non-zero eigenvalues which include six corresponding to libration of the whole subset with respect to its environment: “translation” and “rotation” of the subset as a whole are not free or separable motions but are coupled with the internal vibrational degrees of freedom. This chapter questions the nature of applying the cutoff procedure to environmentally-coupled systems, and the reliability of results issued from the calculations of isotope effects. The use of the cutoff procedure is indeed important in terms of reducing the complexity of Hessian-based calculations, however to what degree can the cutoff procedure be applied in order to reduce computational expense, yet obtain reliable results? Previous studies on the cutoff procedure as applied to COMT had considered errors in ln(KIE), with 1-2% errors being considered acceptable between full and cutoff systems. Two-bond cutoffs as previously suggested by Stern and Wolfsberg have errors of 4% in ln(KIE) for a method derived

from the entire frequency spectrum, and 12% for the so-called translation-rotation method, described herein. The restrictive nature of the two-bond cutoff was therefore suggested to be too extreme for reliable KIE computation, however to what extent is this also true for the model system considered here, and for more compacted and expanded cage structures?

An approximate way to calculate a KIE for a subset of N_s atoms (within a larger environment of $N - N_s$ atoms) is project out the six librational degrees of freedom and to treat them instead as translations and rotations, as mentioned previously.[224] The introduction to this chapter consists of work published in the context of *Kinetic Isotope Effects*, a book chapter written by Prof. Ian H. Williams and PBW in *Simulating Enzyme Reactivity,* and edited by Moliner and Tuñon.

In practice, the $3N_s$ x $3N_s$ Hessian in Cartesian coordinates is first transformed into a $(3N_s - 6)$ x $(3N_s - 6)$ Hessian in a set of internal coordinates, and then back-transformed, whereupon $3N_s - 6$ pure vibrational eigenvalues and 6 zero eigenvalues for translational and rotational motion are obtained.

This procedure is equivalent to treating the subset as if it were a freely translating and freely rotating cluster isolated from its environment but whose internal vibrations are governed by force constants whose values are determined by the influence of the environment. The subset KIEs may then be determined by replacing N by N_s in Equations 15 to 20. However, is there a form, analogous to the Bigeleisen equation, which expresses the KIE in terms only of vibrational frequencies for a subset? The product of the $3N$ eigenvalues of the vibrational secular matrix is equal (Equation 32) to the product of determinants of the Wilson **F** and **G** matrices for potential energy and kinetic energy,[23] respectively, which in turn is equal to the product of determinants of the potential energy matrix **V** and kinetic energy matrix **T** in mass-weighted Cartesian coordinates.[175] There is a problem when there are zero eigenvalues, as for translation and rotation, since then Equation 32 is not valid; the solution is the Teller-Redlich product rule for molecules whose $3N$ degrees of freedom are separable into 3 translations, 3 rotations and $3N - 6$ vibrations.

$$\prod_i^{3N} \frac{\nu_i'}{\nu_i} = \left[\frac{|\mathbf{V}'|}{|\mathbf{V}|}\right]^{1/2} = \left[\prod_j^{N} \frac{m_j}{m_j'}\right]^{3/2}$$

$$\prod_i^{3N} \lambda_i = |\mathbf{F}||\mathbf{G}| = |\mathbf{V}||\mathbf{T}| \qquad (32)$$

However, using the Cartesian formulation of the vibrational problem, it is easy to note that since the frequencies are proportional to the square roots of the vibrational eigenvalues, and **T** is the unit matrix, the product of the ratios of frequencies for isotopologues with the same Hessian (within the Born-Oppenheimer approximation) reduces simply (Equation 33) to the ratio of products of atomic masses used to mass-weight the Hessian prior to diagonalisation. Furthermore, the atomic masses are the same for both isotopologues, except for the isotopically substituted atom(s) alone.

$$f_{\mathrm{RS}} = \left[\prod_{\mathrm{i}}^{3N_{\mathrm{s}}} \underbrace{\frac{\nu_{\mathrm{i}}'}{\nu_{\mathrm{i}}}}_{\mathrm{VP}} \underbrace{\frac{[1-\exp(-u_{\mathrm{i}})]}{[1-\exp(-u_{\mathrm{i}}')]}}_{\mathrm{EX}} \underbrace{\frac{\exp(u_{\mathrm{i}}/2)}{\exp(u_{\mathrm{i}}'/2)}}_{\mathrm{ZP}}\right]_{\mathrm{RS}} = \left[\prod_{\mathrm{j}}^{N_{\mathrm{s}}} \frac{m_{\mathrm{j}}}{m_{\mathrm{j}}'}\right]^{3/2} \prod_{\mathrm{i}}^{3N_{\mathrm{s}}} \left[\frac{\sinh(u_{\mathrm{i}}/2)}{\sinh(u_{\mathrm{i}}'/2)}\right]_{\mathrm{RS}} \tag{33}$$

With a subset Hessian there are, in general, $3N_{\mathrm{s}}$ non-zero vibrational frequencies and so there is no difficulty in evaluating the product on the left-hand side of Equation 33 over $3N_{\mathrm{s}}$ ratios ν_i'/ν_i for an isotopologous pair. Thus there is no need to invoke the Teller-Redlich product rule when a subset Hessian is used. The product of the isotopic ratios for all $3N_{\mathrm{s}}$ frequencies is equal to the three-halves power of the product of the ratios of atomic masses for the specific isotopic substitution. In the context of a typical calculation, satisfaction of this rule requires nothing more than that the matrix diagonalization is performed correctly.[14]

Inspection of either Equation 19 or 20 shows that a KIE consists of a product of RS factors divided by a product of TS factors, which may be re-written in the form of Equation 34.

$$k \,/\, k' = f_{\mathrm{RS}} \,/\, f_{\mathrm{TS}} \tag{34}$$

The terms f_{RS} and f_{TS} are isotopic partition function ratios (IPFRs), as first introduced by Bigeleisen.[9] Whereas an isotopic fractionation factor is defined as the ratio of a heavier isotope relative to a lighter isotope at a particular site with respect to the corresponding ratio at a site in a reference molecule, the definition of an IPFR does not involve any external reference. The logarithm of an IPFR is proportional to the free-energy change due to isotopic substitution at a particular site; an IPFR defined as "heavy"/"light" always has a value >1, becoming larger at lower temperatures but tending towards unity as the temperature increases.

In the light of these considerations, an alternative way to calculate KIEs and IPFRs for subsets of atoms is simply to replace the MMI factor in Equation 15 by extending the products for EXC and the summations for ZPE over $3N_s$ (for RS) or 3 N_s - 1 (for TS) vibrations and librations, as in Equations 35 and 36, in which the contributing factors VP, EX and ZP are identified for clarity. Note that VP may be replaced by the MI factor, which differs from unity only for isotopically substituted atoms within the subset N_s; however, these factors cancel from the KIE (Equation 37) since the same isotopic substitution applies to both RS and TS. Note also that the one-dimensional Bell quantum correction (QC) for an inverted parabola is included in the TS IPFR. Within the approximation of separability, this also corresponds to a free energy change due to isotopic substitution in the TS: for a vibrational mode with a real frequency, a heavier isotope lowers the free energy, but for vibrational mode with an imaginary frequency, a heavier isotope raises the free energy.

$$f_{TS} = \frac{\nu_{\ddagger}'}{\nu_{\ddagger}} \left[\prod_{i}^{3N_s-1} \underbrace{\frac{\nu_i'}{\nu_i}}_{VP} \underbrace{\frac{[1-\exp(-u_i)]}{[1-\exp(-u_i')]}}_{EX} \underbrace{\frac{\exp(u_i/2)}{\exp(u_i'/2)}}_{ZP} \right]_{TS} \underbrace{\frac{\nu_{\ddagger}'}{\nu_{\ddagger}} \frac{\sin(u_{\ddagger}/2)}{\sin(u_{\ddagger}'/2)}}_{QC} \tag{35}$$

$$= \left[\prod_{j}^{N_s} \frac{m_j}{m_j'} \right]^{3/2} \prod_{i}^{3N_s-1} \left[\frac{\sinh(u_i/2)}{\sinh(u_i'/2)} \right]_{TS} \frac{\nu_{\ddagger}'}{\nu_{\ddagger}} \frac{\sin(u_{\ddagger}/2)}{\sin(u_{\ddagger}'/2)} \tag{36}$$

$$k/k' = \frac{\prod_{i}^{3N_s} \left[\frac{\sinh(u_i/2)}{\sinh(u_i'/2)} \right]_{RS}}{\prod_{i}^{3N_s-1} \left[\frac{\sinh(u_i/2)}{\sinh(u_i'/2)} \right]_{TS} \frac{\nu_{\ddagger}'}{\nu_{\ddagger}} \frac{\sin(u_{\ddagger}/2)}{\sin(u_{\ddagger}'/2)}} \tag{37}$$

The use of Equation 37 to calculate a KIE for a subset of atoms within a supramolecular system means that the isotopic sensitivity of all $3N_s$ degrees of freedom is considered, including that of the librational modes that arise from coupling between the subset and its environment. As mentioned above, to project out these contributions to the subset Hessian is equivalent to treating the subset atoms as a cluster in a vacuum. Of course, the supramolecular environment (protein or solvent) affects the values of the Hessian elements corresponding to the $3N_s - 6$ degrees of freedom, which are not the same as they would be from a strictly gas-phase calculation for the subset atoms in isolation, but exclusion of the six external degrees of freedom implies a loss of information in regard to coupling between the subset and its environment. The lowest six of the $3N_s$ frequencies from a subset Hessian calculation for either a RS or TS species typically have values <50 cm^{-1} (a threshold value used by some authors to exclude frequencies from KIE calculations in ISOEFF98),[17] but it is not possible to associate these with the six librations without inspection of the corresponding normal modes, since there is coupling between the internal and external degrees of freedom.

Simply to exclude the lowest six (real) frequencies from a subset KIE calculation would appear to be an arbitrary and risky action for the sake of keeping $3N_s - 6$ frequencies for use with the Bigeleisen equation when this is both invalid and unnecessary for a subset embedded within a supramolecular environment.[105] Pragmatically it may be justifiable provided that these excluded frequencies are isotopically insensitive, but it should not be assumed that this is always the case.

If a Hessian is computed for an atomic subset whose geometry has been relaxed to a local energy minimum (for RS) or saddle point (for TS) within a frozen environment, there may remain a small number of very low-valued imaginary frequencies. The question then arises as to how these should be treated in an IPFR or KIE calculation. Should they simply be excluded, or should they be treated as if they were real frequencies? An argument from experience in favour of adopting the latter option is that it leads to smaller standard deviations from the mean in averages taken over samples of different thermally accessible conformations. Also, exclusion of low-valued but nonetheless isotopically sensitive imaginary frequencies can lead to physically unrealistic IPFRs with values < 1.[224]

From an operational point of view, it is easier to retain all $3N_s$ frequencies in a subset Hessian KIE calculation than it is to perform the projection required to obtain $3N_s$ – 6 pure vibrations and 6 translations and rotations. Moreover, in a study of QM/MM subset Hessians for explicitly solvated AdoMet, it was found that inclusion of all frequencies (in Equation 37) led to significantly smaller errors in KIEs calculated within models with increasingly severe cut-offs than were involved in the more-traditional method (Equation 36) which relied on the Teller-Redlich product rule.[14] Our recommendation is to include all $3N_s$ frequencies and to treat small imaginary frequencies as if they were real. Furthermore, for reactions not involving making or breaking of bonds to hydrogen atoms, the routine application of the one-dimensional approximate quantum-correction to the reaction coordinate frequency is recommended. This applies within the IPFR expression for a TS (Equation 36) and in evaluation of the KIE (Equation 37); in this way the isotopic sensitivity of all $3N_s$ frequencies is considered for both RS and TS.

Truhlar, Gao and co-workers described a method for inclusion of QM vibrational energy into a PMF,[225] which is a necessary part of their EA-VTST procedure.[68, 89, 225, 226] At any point along a reaction path, the reaction coordinate is projected out (to first order) of the Hessian leading to $3N_s$ – 1 generalised normal modes and frequencies. They noted that, for a gas phase molecule, translations and rotations would also be projected out, but that in a condensed phase these are converted into librations, and so the projection does not include these degrees of freedom. However, they also commented that the six lowest-frequency modes of the generalised TS corresponded to very low-frequency librational motion, for which the harmonic approximation was inapplicable and, noting that the quantum correction to low frequency modes was expected to be small, stated that these modes were omitted. Thus the summation of (harmonic) zero-point energy and vibrational partition-function contributions to the free energy was 'calculated over the $3N_s$ – 7 highest-frequency modes, not the $3N_s$ – 1 modes.'[225] While it is true that the librational modes are likely to be anharmonic in nature, and that the harmonic expressions for zero-point energy and the vibrational partition function are probably inaccurate, our view, however, is that it is still better to include the isotopic sensitivity of these modes, notwithstanding the limitations of the harmonic approximation, than to exclude them altogether.

6.2 The Cut-off rules

The motivation underlying the original cut-off rule proposed by Stern and Wolfsberg[15] was to make KIE calculations tractable in regard both to computational resource and availability of force-constant data for RS and TS structures. In view of the fact that they considered only covalently-bound molecules, the rule was stated in terms of how many covalent bonds a particular atom is removed from the site of isotopic substitution.[227] Simulations of chemical reactivity in enzyme active sites or in solution (or other condensed media) inevitably involve consideration of non-covalent interactions. The range of QM and QM/MM applications surveyed above testifies that neither lack of computer power nor the ability to compute a Hessian is now a limiting factor. But a question remains: what size does a Hessian need to be in order to provide reliable estimates for IPFRs, and therefore KIEs? How should the cut-off rule be re-framed for KIE calculations in the supramolecular age?

In a QM/MM study of cut-offs on calculated KIEs for methyl transfer AdoMet to catecholate in water,[224] it was found that, for more-severe cut-offs, significantly smaller errors were obtained when the six librational degrees of freedom were treated as harmonic vibrations than as being projected out as translations and rotations. Errors were evaluated by comparison of the natural logarithm of the KIE for a cut-off Hessian against that obtained from a large Hessian, in this case for 324 atoms of AdoMet and 87 water molecules. The QM region (AM1) contained 63 atoms of AdoMet and the MM region contained about 500 waters (TIP3P). Cut-off Hessians were obtained by deletion of appropriate rows and columns from the full Hessian; the values of retained elements in each Hessian remained constant throughout. Exclusion of all solvent water molecules from the Hessian yielded β-$^{2}H_3$, α-^{13}C and nucleofuge ^{34}S KIEs with equally small errors using either the “all frequencies” method or the projection method, but the latter method (using only $3N_s - 6$ frequencies) gave a much larger error for the nucleophile ^{18}O KIE. The standard 2-bond cut-off rule performed significantly better with the all $3N_s$ frequencies method for all isotopic substitutions except α-^{13}C, for which the error was the same with both methods. Larger errors were obtained with both methods when the one-dimensional Bell tunnelling correction was included, because the imaginary TS reaction-coordinate frequencies were surprisingly sensitive to the size of the subset Hessian. It was concluded that, to ensure that the error in ln(KIE) remained less than 1% (or 2% for the quantum-corrected KIE) for all four isotopic substitutions considered, it

was necessary to use a less-restrictive cut-off procedure which retained all covalently-bonded atoms to a distance of at least 3-bonds together with solvent water molecules whose O atom was less than 4.5 Å from any isotopically substituted atom. To the extent that KIEs may be treated to a satisfactory degree of approximation by means of a cut-off procedure, it implies that these ratios of rate constants are essentially local properties of the system, reflecting changes in curvature of the potential energy surface in the immediate vicinity of the position(s) of isotopic substitution. However, the surprising degree of sensitivity of KIE values to the extent of cut-off implies that these are not entirely local properties of the system. In particular, it seems that a cut-off should not cause significant change in the nature of an isotopically sensitive reaction-coordinate vibrational mode: exclusions of atoms whose motions are coupled to the reaction coordinate may lead to error. These findings have important practical implications for reliable computational simulation of KIEs for large systems in solution or within an enzyme active site.

To obtain reliable KIEs requires not only an appropriate choice of QM or QM/MM method but also an adequate description of the environment in regard to both its potential energy and kinetic energy influence upon the positions of isotopic substitution.[224]

In QM/MM applications, the atoms included in the subset Hessian have commonly been the same as those in the QM region. However, the selection criteria for inclusion of an atom in the QM region are not necessarily the same as those for inclusion in the Hessian. A QM region may be chosen so as to avoid the need for the QM/MM boundary to cut through covalent bonds. However, this issue does not affect the Hessian: there is no reason why even just a single atom of a group connected, either covalently or non-covalently, within two bonds distance from a site of isotopic substitution cannot be included in a subset Hessian. The size and shape of the atomic subsets included in the Hessian and in the QM region do not need to be the same; better KIE results may be obtained using QM/MM Hessians computed for roughly spherical subsets, centred on the site of isotopic substitution, whose size is determined by a “cut-off” radius. The quality of the data contained within a computed subset Hessian is as important as the quantity: reliable KIE calculations for enzyme-catalysed reactions require an adequate description of the influence of the (generally inhomogeneous) environment surrounding the subset, which in turn may require a larger QM region than has typically been used in the past.

An early technique in molecular simulation involved the implementation of the cutoff approximation, discussed at length in Chapter 1. This was based on simplifying calculations by including only atoms which were a specified number of covalent bonds away from the site of isotopic substitution, in the calculation of the Hessian. This was advantageous in early computational chemistry due to the limitations of the hardware. Now however, it is possible to deal with larger and larger systems.

The issue that remains is the computational expense of calculating the properties of enzymic and supramolecular systems. It is therefore possible to use the cutoff approximation in terms of reliable data as opposed to simplicity. At which point does a cutoff treatment become too strict, and moreover, when is it possible to obtain results close to those for full systems, to avoid unnecessarily complex computations?

Following on from the effect of solvent, which has been well-documented in previous chapters, and the impact of anharmonic corrections; incorporating the cutoff procedure into our work was the next logical step in the story. This was therefore applied to the cage system, as well as a test on the isotope effects from QM/MM calculations.

6.3 The Cutoff Procedure: Applications to the *cage* system

This section of the chapter presents recent results of the testing procedure surrounding the cutoff approximation. A constrained cluster consisting of the methyl cation surrounded by 5 water molecules in axial and equatorial arrangements, as well as structures output from QM/MM simulations were used to develop our understanding of the explicit effects of the environment on isotope effects calculated from a subset of the atoms, as compared to the full system. It is possible to conclude that caution must be used when applying the cutoff procedure to systems where environmental effects are intrinsically coupled to the atoms of interest. In particular, care must be taken when manipulating Hessian elements directly affecting the reaction coordinate for computation of KIEs.

As first discussed in Chapter 1, and now having been developed throughout Chapter 6, it is important to further our understanding of the limits of the cutoff procedure for application to large systems. Williams' publication where numerous iterations of the cutoff protocols were applied to an AdoMet system has been discussed in earlier;

a similar investigation follows here in a model context, and in Chapter 7 with recently-obtained QM/MM structures and Hessian matrices.

All cage electronic structure calculations were carried out using the Gaussian 09 code,[174] with the B3LYP functional and aug-cc-PVDZ basis set. The constrained methyl cage (Fig. 12) was constructed using a z-matrix and C – O distances then varied between 2.04 and 5 Å axially, and 3.0 and 5.0 Å equatorially. Transition structures for computation of the KIEs were obtained using the Berny algorithm, with reactant and product species for the EIEs being optimised to minima.

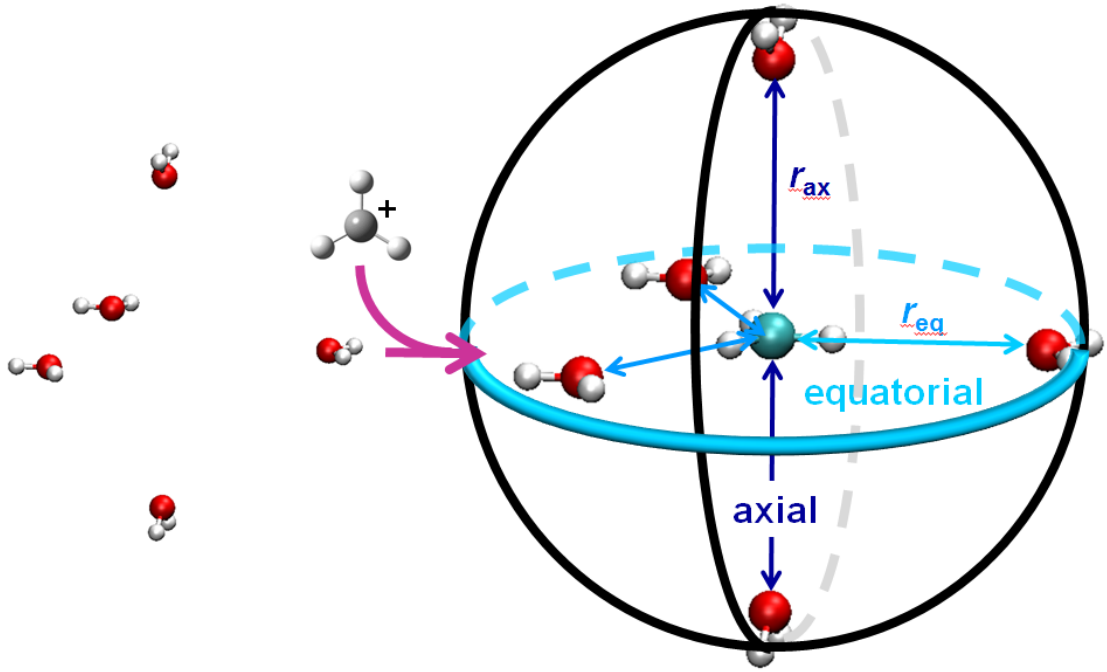


Figure 12. Illustration of the isotope effect on going from solvent cage to methyl cation in a constrained environment. Figure extracted from Ref 3.

Isotope effects were obtained using the suite of vibrational analysis programs discussed in Chapter 2. In order to consider the effect of the environmental coupling on the methyl cation, the superheavy cage approach was again used (Chapter 5), whereby the hydrogen atoms on the water molecules were given masses of 999 a.m.u as a substitute for a large protein structure with the methyl cation in an active site location. The isotope effects for the full system, and a number of subsets were computed in order to gauge the applicability of the cutoff procedure for such a constrained test geometry. Subsets considered, each contained the methyl cation as

well as: axial water oxygen atoms, equatorial water oxygen atoms, and all oxygen atoms. These reductions to the Hessian took place using the procedure described above, with only the data corresponding to the selected atoms being taken from the full system Hessian.

6.3.1 Results and Discussion

The elements of the Hessian relating to the methyl cation itself were selected, and the procedure for calculation of the isotope effect repeated. The results for r_{eq}=3, r_{ax}=3 are shown in Figure 14. These show the KIEs for the full system, compared to those output from calculations using only the Hessian elements for the methyl cation extracted from the full system. These elements therefore retain (some of) the environmental effects relating to the full system, whilst explicitly describing the CH_3 group itself.

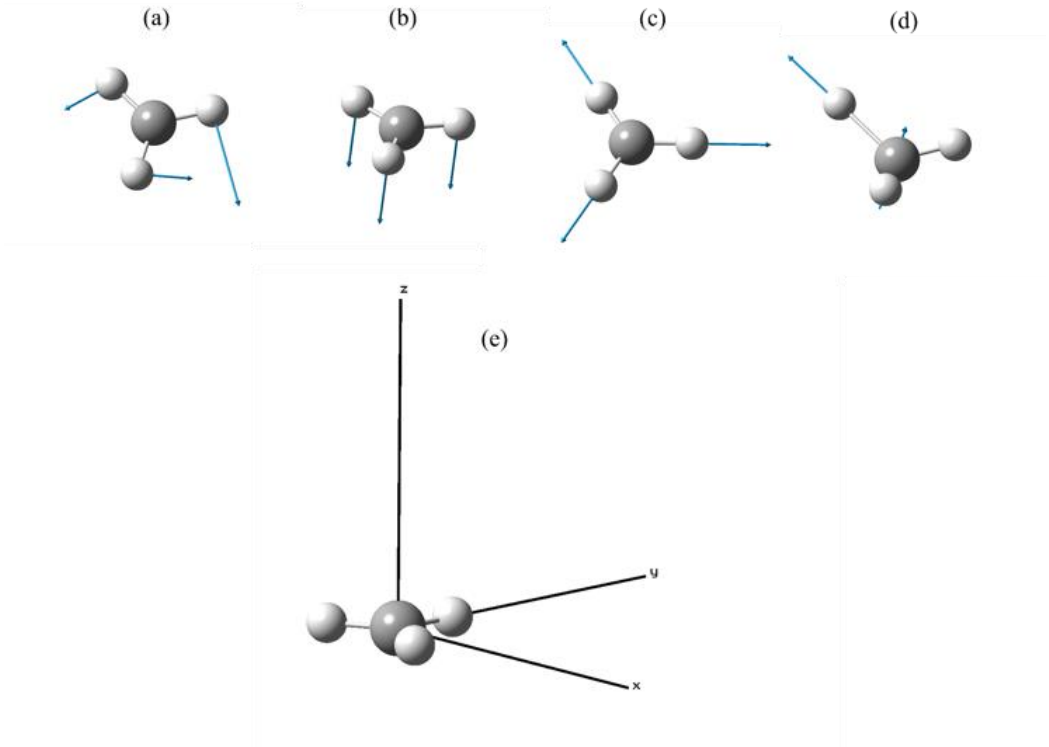


Figure 13(a) to (e). Diagrammatic representation of the 12 modes of the methyl cation. (a) describes the degenerate bending mode, (b) the out of plane bend, (c) the symmetric stretching mode, (d) the degenerate asymmetric stretching mode, and (e) shows the Cartesian axes. (e) can be considered as depicting the three translation and rotational motions of the external modes, along/about each of the Cartesian

axes. Indeed, this represents one translation along each of the axes, and one rotation about these.

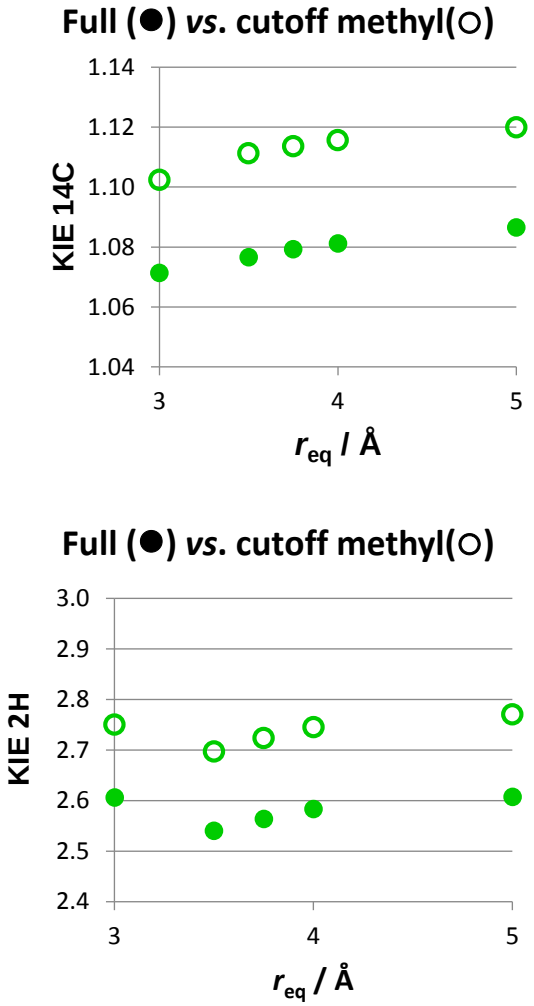


Figure 14. Full and cutoff 1º ^{14}C and 2º α-D KIEs for the solvated cluster, and cutoff methyl cation, respectively.

The trends are reproduced remarkably well, however with a difference in the magnitudes. Carrying out individual frequency analyses to determine the effect of each mode on the total isotope effect, it has been determined that the external modes (translations and rotations) of the methyl cation, are responsible for the trends observed in the KIEs (Figure 15).

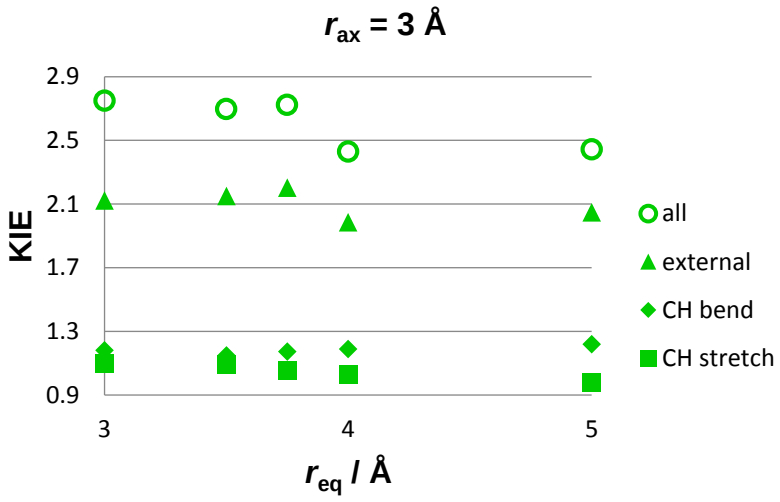


Figure 15. Cutoff 2° α-D KIEs at 3 Å for the methyl group transfer in a constrained cage, categorised by type of vibrational mode.

The influence of the CH bends and stretches contributes to the overall magnitude of the KIE, however it is clear that the external modes dictate the trend observed. This is reproduced for each of the axial distances, and is suggestive of a gradual decline in the applicability of the cutoff procedure as the axial distances increase. Unsurprisingly, this appears to lead to a poorer description of the system in general as the reaction coordinate mode (translation in the *z* vector) is no longer well-described at large distances and therefore it is notable that the exploded S_N1-like structures begin to lack imaginary frequencies incorporating reaction coordinate motion and can therefore be considered EIEs.

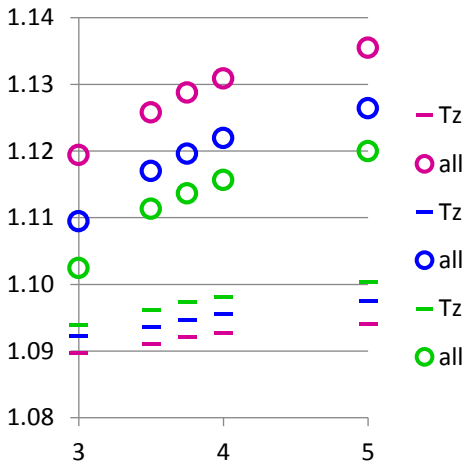


Figure 16. 1° ^{14}C cutoff KIEs for methyl cation transfer between two water nucleophiles, with trends of the *z*-axis translation component for 2.04 (purple), 2.5 (blue) and 3.0 (green) axial water nucleophile distances included as tabs.

The 1° ^{14}C cutoff KIEs are shown in Figure 16. It can be seen that the external modes dominate the trends observed in total KIE, with the internal degrees of freedom being responsible for only part of the magnitude but not general variation observed with an increase in equatorial distance. This is confirmed by the controlling influence of the T_z mode, or reaction coordinate frequency (RCF), which itself governs the trend in the external modes. This is of particular importance for the 1° ^{14}C KIE, which will by default, be affected to a greater degree by erroneous T_z components than the 2° α -D KIEs.

Table 18. Contribution of the T_z mode (translation in the *z* direction) to the total ^{14}C KIE for methyl transfer with a constrained cage of 5 water molecules.

		cutoff					
		3	3.5	3.75	4	5	∞
2	Tz	1.090	1.091	1.092	1.093	1.094	1.095
2	all	1.119	1.126	1.129	1.131	1.135	1.140
2.5	Tz	1.092	1.094	1.095	1.096	1.098	1.099
2.5	all	1.109	1.117	1.120	1.122	1.126	1.127
3	Tz	1.094	1.096	1.097	1.098	1.100	1.102
3	all	1.102	1.111	1.114	1.116	1.120	1.118
4	Tz	1.098	1.033	1.101	1.096	1.103	1.101
4	all	1.103	1.043	1.112	1.104	1.102	1.115
5	Tz	1.034	1.035	1.035	1.036	1.104	1.101
5	all	1.041	1.046	1.046	1.047	1.117	1.114

The conclusion that can therefore be drawn for all the above, rather drastic methyl cation cutoffs, is that additional atoms must be included in order to better describe the environment, and in particular, the reaction coordinate frequency.

In order to better understand the effects observed, and to attempt to obtain more reliable Hessians, the degree of cutoff was relaxed to produce two independent sets of results. Firstly, the oxygen atoms in the axial water molecules are included in addition to the methyl cation (referred to as *axialO*), and secondly, also including the oxygen atoms from the equatorial water molecules (referred to as *ax/eq*). This would provide a gradual sense of the environmental impact of cutoff tightness and Hessian reliability affecting the computed KIEs, while relating to the *x*-bond cutoff methodology of Wolfsberg and Stern.

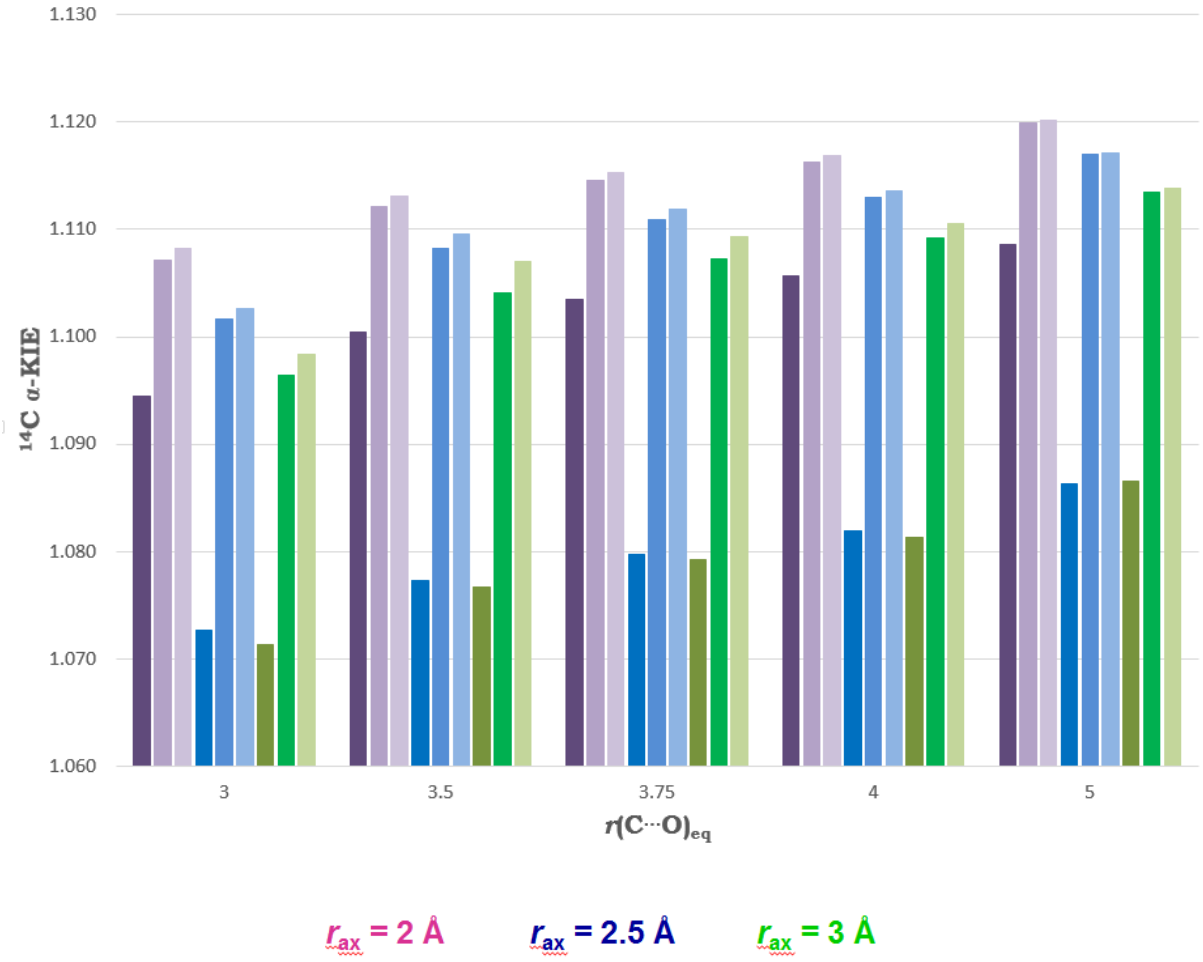


Figure 17. The effect of including the full Hessian in IPFR evaluation (dark respective colours), as compared to progressively smaller Hessians, representing subsets of the system, where lighter represents the methyl cation, and the axial and equatorial oxygen atoms, excluding their hydrogen atoms, and lightest respective colour corresponds solely to elements retained in the Hessian for the methyl cation and both of the axial oxygen atoms.

Including only the axial oxygen atoms within the cutoff procedure begins to improve the accuracy of the cutoff results obtained when compared to the methyl cation cutoff, by showing a progressive trend in the KIEs. This is also characterised and confirmed by the translation in the *z* vector containing the appropriate imaginary frequency for the majority of cutoff structures. Again, the magnitudes of the cutoff KIEs differ from the full treatment; however the trend appears to be somewhat more respected than in the more drastic cutoff.

Including the axial *and* equatorial oxygen atoms (Figure 17) again has an increasingly positive effect on the proximity of the cutoff results to the full treatment. Not at all surprising as the hydrogen atoms remain the only entities still removed from the cutoff procedure, however significantly better agreement between the trends in the full system and those in this two-bond cutoff are now observed, although some magnitude-related irregularities still remain.

Using Figure 18 to represent different cutoff levels, Table 19 details the effective errors in the natural logarithms of the KIEs; methodology described earlier in this chapter.

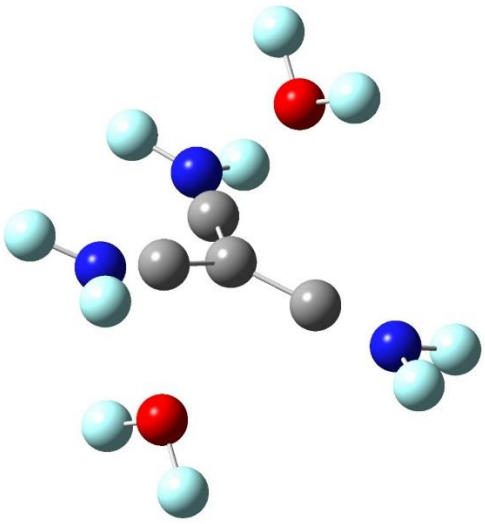


Figure 18. Reference used to interpret Table 19. The grey atoms represent the methyl cation, or *cutoff system 4 i.e.* the most drastic cutoff. The red atoms (axial oxygen atoms) represent the next level of cutoff, in combination with level 4, *i.e.* methyl cation + axial oxygen atoms = *cutoff level 3*. The blue atoms (equatorial oxygen atoms) show the second level of the structure, or *cutoff level 2*, in combination with the previous levels, making: equatorial oxygen atoms, axial oxygen atoms, and methyl cation. Finally, what is referred to as *level 1* in Table 19 represents the full, non-cutoff system, with the inclusion of all previous levels, as well as the water hydrogen atoms.

Table 19. Relative errors in *ln*(α-^{14}C KIE) for cutoff cage structures relative to full system. r_{ax} and r_{eq} represent the axial C···O and equatorial C-H···O distances in Angstrom, respectively. Cutoff level indicates the extent to which the system Hessian is reduced in size with 1 referring to only the contributions to the Hessian of hydrogen atoms on water molecules being removed, to 4 all contributions to the Hessian from atoms other than the methyl cation.

r_{ax}	r_{eq}	cutoff level	KIE	ln(KIE)	% difference in ln(KIE)
2.04	3	1	1.0945	0.0903	
	3.5	1	1.1004	0.0957	
	3.75	1	1.1035	0.0985	
	4	1	1.1057	0.1005	
	5	1	1.1086	0.1031	
	3	2	1.1071	0.1018	12.7
	3.5	2	1.1121	0.1062	11.0
	3.75	2	1.1145	0.1085	10.1
	4	2	1.1163	0.1100	9.5
	5	2	1.1200	0.1133	9.9
	3	3	1.1083	0.1028	13.8
	3.5	3	1.1131	0.1071	11.9
	3.75	3	1.1153	0.1091	10.8
	4	3	1.1168	0.1105	9.9
	5	3	1.1201	0.1134	10.0
	3	4	1.2046	0.1861	106.1
	3.5	4	1.1853	0.1700	77.6
	3.75	4	1.1803	0.1658	68.3

	4	4	1.1737	0.1601	59.4
	5	4	1.1640	0.1519	47.3
2.5	3	1	1.0727	0.0701	
	3.5	1	1.0773	0.0744	
	3.75	1	1.0797	0.0767	
	4	1	1.0820	0.0788	
	5	1	1.0863	0.0828	
	3	2	1.1016	0.0968	37.9
	3.5	2	1.1082	0.1028	38.1
	3.75	2	1.1109	0.1052	37.1
	4	2	1.1130	0.1070	35.9
	5	2	1.1169	0.1106	33.6
	3	3	1.1026	0.0976	39.2
	3.5	3	1.1096	0.1040	39.8
	3.75	3	1.1119	0.1060	38.2
	4	3	1.1136	0.1076	36.6
	5	3	1.1172	0.1108	33.9
	3	4	2.1814	0.1142	62.8
	3.5	4	2.1112	0.1172	57.4
	3.75	4	2.1130	0.1202	56.7
	4	4	2.1338	0.1232	56.4
	5	4	2.1531	0.1262	52.4

3	3	1	1.0713	0.0689	
	3.5	1	1.0766	0.0738	
	3.75	1	1.0793	0.0763	
	4	1	1.0813	0.0781	
	5	1	1.0866	0.0830	
	3	2	1.0964	0.0920	33.5
	3.5	2	1.1040	0.0990	34.1
	3.75	2	1.1072	0.1018	33.5
	4	2	1.1092	0.1036	32.6
	5	2	1.1134	0.1074	29.4
	3	3	1.0983	0.0938	36.1
	3.5	3	1.1070	0.1017	37.7
	3.75	3	1.1093	0.1037	36.0
	4	3	1.1106	0.1049	34.2
	5	3	1.1138	0.1078	29.8
	3	4	2.6965	0.9919	1339.3
	3.5	4	2.6990	0.9929	1245.0
	3.75	4	2.7297	1.0042	1216.6
	4	4	2.6939	0.9910	1168.2
	5	4	2.7111	0.9973	1101.5

Table 19 is arranged first by axial separation of the carbon atom of the methyl cation, followed by equatorial carbon-oxygen distance (all in Angstroms). Each of the *cutoff levels* indicates a further reduction in the size of the Hessian, with 1 referring to the full system, and 4 the methyl cation.

A number of conclusions relating to the implementation of the cutoff procedure, can be drawn from the data presented in Table 19. As first described earlier in this chapter, Stern and Wolfsberg suggested a two-bond cutoff as the limit for reliable estimation of isotope effects. With the data in Table 19 corresponding to the effect of ^{12}C substitution by ^{14}C, a two-bond cutoff of this model system would amount to the deletion of all rows and columns corresponding to the water hydrogen atoms being deleted from the original Hessian matrix of the system.

The purpose of cutoff levels 2 and 3 is to consider the effect of environmental interactions outside the reaction coordinate, on the reliability of IPFRs obtained for a system. Indeed, cutoff level 2 represents the Stern-Wolfsberg traditional two-bond cutoff in all directions, whereas level 3 considers this solely in terms of the reaction coordinate, removing the elements of the Hessian corresponding to the equatorial, hydrogen-bonding water molecules.

The data presented provides reliable estimation of the impact of so-called *perpendicular* environmental effects on KIEs, but also the degree to which the cutoff approximation can be employed in terms of simplifying calculations and post-processing of data, particularly for large systems.

As was first suggested in Chapter 1, the natural logarithm of the KIE is proportional to the free energy change associated with isotopic substitution. It therefore provides a reasonable metric for considering error based on differences in the value of *ln*(KIE) for different sizes of subset Hessians.

It is therefore useful to estimate the reliability of the 1^{o}-^{14}C KIEs in Table 19, based on the % error in the *ln*(KIE) for a particular level of cutoff, and geometry of structure. The structures with r_{eq} = 3.0 Å for r_{ax} = 2.04, 2.525, and 3.0 Å variations follow a generally monotonic trend in error in *ln*(KIE) as cutoff level increase (becomes more severe). This is particularly evident from the degeneration of the isotope effects at cutoff level 4, where only elements relating to the methyl cation remain in the subset Hessian.

Indeed, for r_{eq}=3.0Å,r_{ax}=2.04Å, the % error in *ln*(KIE) as compared to that of the full Hessian-based KIE, is of 106%, rising to more than 1000% for the symmetric r_{eq}=3.0Å,r_{ax}=3.0Å system. This can be explained by the discrepancies inherent in the description of the reaction coordinate mode for methyl cation cutoffs. As was demonstrated in Figure 16, the translation in the *z* vector is responsible for the shape of the trends in KIE observed, particularly for the α-^{14}C effect. It is therefore important for this mode to be accurately described, in order for reliable IPFRs to be calculated. Removal of the environmental element intrinsically affecting motion in the reaction coordinate results in anomalous description of the mode by remaining Hessian elements, and therefore detrimentally affecting the value of the isotope effect obtained.

Furthermore, for the r_{eq}=3.0Å,r_{ax}=3.0Å and in particular other structures where r_{ax} >3.0Å, the reaction coordinate frequency is not necessarily reproduced to a reasonable magnitude, representing small frequencies often associated with spurious external motions. This is of particular importance, when other codes for calculation of IPFRs omit frequencies with magnitudes <50cm^{-1}, with the possibility of these further reducing the reliability of estimated isotope effects.[17]

Generally, as the level of cutoff goes from 1 through to 4, each sequence of subset Hessian results in increasing errors in *ln*(KIE) as compared to the full system. Indeed, for the r_{eq}=3.0Å,r_{ax}=2.04Å system, the error in *ln*(KIE) increases from ~13% at cutoff level 2, to 14% at level 3, and as previously shown, 106% at level 4.

Moreover, this effect of error is replicated in r_{eq}=3.0Å,r_{ax}=2.525Å, where level 2, including all oxygen atoms produces errors in *ln*(KIE) of ~38%, going to ~39% when including only axial oxygen atoms at level 3, and finally ~63% error for the subset Hessian corresponding to the methyl cation at level 4.

In the final page of Table 20, the r_{eq}=3.0Å,r_{ax}=3.0Å results show again the same increase in error on going from the full Hessian, to each of the cutoff systems. The error for the level 2 of the subset Hessian is ~34%, rising to ~36% at level 3, and the alarming >1000% at level 4.

With increasing values of r_{ax} resulting in degeneration of KIEs obtained from subset Hessians, increases in r_{eq} also affect the KIE although resulting in a decrease in error. Indeed, considering the r_{ax}=2.04Å series, for the cutoff level of 2, where solely the water hydrogen atoms are removed, the error in *ln*(KIE) goes from ~13% at r_{eq}=3.0Å, to 11% at 3.5Å, and further down to 10% and 9% for further increases in r_{eq}.

Increasing the severity of the cutoff procedure to a subsequent level of course has the effect of increasing the overall error as described above, although this decreases with r_{eq} again. The error in *ln*(KIE) for the methyl cation cutoff at r_{ax}=2.04Å is again inversely proportional to r_{eq}, with the most condensed structure at r_{eq}=3.0Å carrying an associated error in *ln*(KIE) of more than 100%, decreasing to ~78% when r_{eq}=3.5Å.

Furthermore, this increase in error for more severe cutoff levels, with increasing r_{eq} is replicated for each value of r_{ax}, also. It is reasonable to assume that removing elements of the Hessian relating to a full environmentally-coupled system would lead to errors in the results obtained from such a procedure. It is however important to quantify these errors with regards to the size of the subset included within the data processing and characterisation procedures.

Indeed, with an increase of r_{ax} the total error for that distance is increased in comparison to that for the shorted C···O bond length. The total errors for each series of r_{ax} distance and each subset of cutoffs increases with cutoff severity, as described above, but also with r_{ax} (Table 20).

Table 20. Cumulative % errors in *ln*(KIE) for each series of subset cutoffs within the group of structures with each of the three values of r_{ax}. These include each of the five r_{eq} variations for each value of r_{ax}, for each of the cutoff levels 2 through 4.

r_{ax} (Å)	Cumulative % error in ln(KIE)
2.040	468
2.525	656
3.000	6408

Table 20 gives a broad view of the error associated with each series of cutoffs, within each of the r_{ax} distance variations, the r_{eq} distances of course being incremented within each level of cutoff.

Over a total of 15 structures, with error increasing with the level of cutoff, the average error in the systems increases from 31% in *ln*(KIE) for the r_{ax}=2.04Å structures, to 44% at r_{ax}=2.525Å, and finally 427% for r_{ax}=3.0Å. Of course, these are influences significantly by the higher order cutoffs leading to larger errors, however this demonstrates the general increase in error associated with cutoffs as the distance in the axial plane increases.

This proportional increase in cutoff system error with larger r_{ax} distance can be attributed to a number of factors. Firstly, as the distance between the methyl cation and the donor and acceptor groups increases (axial and equatorial water molecules), the system behaves less like a traditional prototype S_N2 transfer, and gains more S_N1 character. Indeed, considering the importance of proximity on direct reactions, as well as environmental factors, it is unsurprising that removal of Hessian elements at larger values of r_{ax} have a more significant effect than at smaller ranges. If an interaction is strong, then removal of the species responsible from the Hessian matrix should have only a marginal effect, as the elements relating to the substituent on which the effect is incident, should continue to exhibit the respective properties dependant on such an effect, as the elements relating to this would have been conserved within the Hessian.

If however an effect is measurably weaker due to increased distance between the moiety responsible for the effect, and the substituent on which the effect is incident, then even careful use of a cutoff procedure will remove the little impact of the effect from the elements of the Hessian, with those responsible for the remaining atoms exhibiting little to no residual properties from this.

It has been previously demonstrated that donor-acceptor distances, particularly in AdoMet-dependant methyl transferases vary between approximately 2 and 2.5Å.[228] In the light of this information, it is therefore unsurprising that the results with the least error associated with *ln*(KIE) correspond to r_{ax}=2.04Å. It is also noteworthy that those results representing cutoff levels of 4, are more prone to error, particularly at larger values of r_{ax}, with the elements for the donor and acceptor molecules being omitted from the Hessian altogether.

Table 21. Frequencies (cm^{-1}) for the methyl cation cutoff structures (level 4) for each of the cage geometries. RS and TS refer to reactant and transition structures, respectively. *ax* and *eq* are *axial* and *equatorial* and refer to the C···O and C-H···O distances, respectively within the constrained cage of five water molecules surround the methyl cation. *T,* and *R* refer to the external degrees of freedom, *translations* and *rotations*, with *x,y,* and *z* referring to the Cartesian axes on which these operate. *Umbr* represents the methyl cation umbrella motion between the three hydrogen atoms, *scis*, the scissor bending motion of C-H bonds, *rock* the rocking motion, and *CH* the C-H stretching motions.

	RS		TS		
ax 2.04	frequencies/cm^{-1}			frequencies/cm^{-1}	
eq 3.0	H_3	D_3		H_3	D_3
Tx+Ry	224.7255	201.0619	RCF (Tz)	-260.4759	-253.7355
Rz	245.9594	174.4126	Tx	284.0782	260.0611
Ty+Rx	299.3508	264.4841	Rz	292.6553	207.3354
Tz	422.8806	407.6927	Ty	341.2214	311.3845
Tx+Ry	1169.278	856.6097	Ry	1025.8699	725.733
Ty+Rx	1189.79	879.0732	Rx	1054.0355	745.602
umbr	1425.6	1039.524	opla	1360.3335	987.7587
scis	1457.491	1056.703	def	1440.8148	1053.862
rock	1487.734	1078.837	def	1462.5536	1070.966
CH	3122.414	2221.113	CH (a1)	3105.1882	2196.544
CH	3281.154	2445.478	CH	3312.6529	2475.52
CH	3284.547	2449.99	CH	3315.8806	2479.206
ax 2.04					
eq 3.5					
Tx+Ry	126.6378	113.0170	RCF (Tz)	-275.6772	-269.3514
Ty+Rx	234.8189	210.5960	Tx	198.6710	181.2821
Rz	263.7385	183.6644	Ty	271.4778	252.8007
Tz	443.4564	427.0260	Rz	287.9209	199.2084
Tx+Ry	1161.8383	852.1158	Ry	1016.5190	719.0950
Ty+Rx	1177.8356	871.3583	Rx	1049.2644	742.2270
umbr	1412.0142	1030.0885	opla	1324.4728	958.8652
scis	1437.8035	1043.6870	def	1415.4512	1037.9812
rock	1467.4503	1065.4699	def	1437.8869	1055.7474
CH	3086.9933	2196.5668	CH (a1)	3084.2807	2181.7546
CH	3248.8368	2422.8985	CH	3300.6856	2468.2696
CH	3252.1050	2427.2452	CH	3301.5142	2470.3292
ax 2.04					
eq 3.75					
Tx+Ry	96.1142	85.6063	RCF (Tz)	-282.8212	-276.6124
Ty+Rx	216.0855	201.1876	Tx	185.1967	168.8237
Rz	235.2043	157.4583	Ty	271.6368	245.4737
Tz	456.6827	439.3959	Rz	239.8620	170.8406
Tx+Ry	1153.6530	847.7768	Ry	1006.2911	711.8598
Ty+Rx	1170.5989	867.7005	Rx	1039.9876	735.6648

umbr	1405.0695	1025.5403	opla	1309.3550	946.9577
scis	1429.6468	1038.0011	def	1405.1953	1031.4860
rock	1459.9517	1060.2455	def	1427.9824	1049.5896
CH	3092.8313	2201.2320	CH (a1)	3094.7029	2189.1263
CH	3252.2854	2425.2023	CH	3311.1235	2476.0077
CH	3256.3371	2430.0279	CH	3312.3740	2478.2908
	RS			TS	
ax 2.04	frequencies/cm^{-1}			frequencies/cm^{-1}	
eq 4.0	H_3	D_3		H_3	D_3
Tx+Ry	84.4960	75.1630	RCF (Tz)	-288.4822	-282.3679
Ty+Rx	183.1940	130.5995	Tx	182.5000	166.2653
Rz	226.0223	197.3034	Ty	268.5691	243.5346
Tz	466.0227	448.1119	Rz	201.3417	142.8226
Tx+Ry	1147.9470	844.7682	Ry	998.3504	706.2463
Ty+Rx	1164.6657	864.7071	Rx	1031.7246	729.8197
umbr	1399.5158	1021.7325	opla	1297.6157	937.7353
scis	1424.0667	1034.1891	def	1397.5402	1026.6342
rock	1454.3955	1056.3710	def	1420.5193	1044.9050
CH	3101.1603	2207.5525	CH (a1)	3106.6802	2197.5994
CH	3258.7154	2429.5526	CH	3323.1991	2484.6759
CH	3263.1574	2434.6130	CH	3323.9406	2486.6304
ax 2.04					
eq 5.0					
Tx+Ry	75.0536	53.2058	RCF (Tz)	-299.9941	-294.1033
Ty+Rx	81.1470	72.0061	Tx	186.1785	169.4455
Rz	224.1245	196.1308	Ty	271.8457	246.6664
Tz	485.4471	466.0977	Rz	90.1941	63.8684
Tx+Ry	1136.9519	839.1812	Ry	982.6714	695.1563
Ty+Rx	1153.7844	859.6715	Rx	1017.1496	719.5098
umbr	1390.7879	1016.0823	opla	1274.2487	919.3830
scis	1415.4130	1028.2830	def	1383.7804	1017.9130
rock	1446.2155	1050.5810	def	1407.2154	1036.6231
CH	3122.5521	2223.6576	CH (a1)	3134.1245	2217.0126
CH	3274.9445	2440.4066	CH	3349.0492	2503.1158
CH	3279.8212	2445.7344	CH	3349.6680	2504.9654
ax 2.525					
eq 3.0					
Tx+Ry	154.2	137.3046	RCF (Tz)	-206.7233	-193.3161
Rz	230.4314	201.8576	Tx	241.2611	221.3503
Ty+Rx	300.7491	213.7146	Rz	263.9511	242.0559
Tz	476.3731	453.5286	Ty	372.3723	263.4194
Tx+Ry	1141.977	842.5011	Ry	592.1761	418.8926
Ty+Rx	1155.92	860.9012	Rx	629.6318	445.3879

umbr	1450.247	1058.385	opla	1391.3711	1052.486
scis	1467.21	1072.104	def	1454.0597	1060.931
rock	1490.288	1078.823	def	1460.9838	1066.291
CH	3059.503	2179.702	CH (a1)	2996.6068	2119.738
CH	3207.039	2390.56	CH	3195.8837	2388.89
CH	3209.212	2394.233	CH	3198.6231	2391.254
	RS			TS	
ax					
2.525	frequencies/cm^{-1}			frequencies/cm^{-1}	
eq 3.5	H_3	D_3		H_3	D_3
Tx+Ry	73.3404	65.1453	RCF (Tz)	-210.7007	-197.2908
Rz	193.2740	168.7202	Tx	138.2645	126.4413
Ty+Rx	262.9713	187.0297	Rz	175.8149	160.6976
Tz	489.8284	465.4108	Ty	316.7476	224.0623
Tx+Ry	1131.7489	835.8279	Ry	593.8690	420.0908
Ty+Rx	1146.7829	855.0037	Rx	626.9574	443.4961
umbr	1433.8226	1046.3752	opla	1344.1624	1015.4583
scis	1448.0449	1061.3792	def	1426.6125	1044.4589
rock	1470.7345	1066.3782	def	1434.2072	1050.3023
CH	3057.0558	2178.5525	CH (a1)	3016.1102	2133.5318
CH	3205.5759	2389.5765	CH	3230.4987	2414.4022
CH	3208.9626	2393.9339	CH	3231.0131	2415.3643
ax					
2.525					
eq 3.75					
Tx+Ry	34.7928	30.8535	RCF (Tz)	-216.4484	-202.7333
Rz	186.0111	154.4878	Tx	119.7925	109.4477
Ty+Rx	225.3581	168.1853	Rz	163.4510	149.2577
Tz	501.1629	475.7873	Ty	268.7039	190.0776
Tx+Ry	1127.0385	833.8024	Ry	574.3511	406.2838
Ty+Rx	1143.0790	853.6653	Rx	610.1597	431.6138
umbr	1428.3448	1041.7113	opla	1325.2666	1000.885
scis	1441.2025	1057.8863	def	1415.7451	1037.666
rock	1464.3164	1061.9951	def	1423.6769	1043.779
CH	3066.8020	2185.8168	CH (a1)	3034.8597	2146.7945
CH	3213.5592	2395.0506	CH	3250.7942	2429.0907
CH	3217.8189	2399.9360	CH	3251.5096	2430.1743
ax					
2.525					
eq 4.0					
Tx+Ry	-33.8154	-29.9448	RCF (Tz)	-221.0820	-207.1263
Rz	174.6419	129.2793	Tx	113.9850	104.0807
Ty+Rx	193.1155	161.4366	Rz	160.5919	146.5544
Tz	511.3564	485.0676	Ty	223.2145	157.9051
Tx+Ry	1123.1621	832.2847	Ry	558.1721	394.8393
Ty+Rx	1139.4463	852.4907	Rx	595.6843	421.3742

umbr	1423.7882	1038.1956	opla	1311.0662	989.9062
scis	1436.4430	1055.2899	def	1407.8986	1032.728
rock	1459.7799	1058.7647	def	1415.9754	1038.979
CH	3076.8496	2193.3063	CH (a1)	3051.8361	2158.8031
CH	3222.0137	2400.8337	CH	3268.2422	2441.6350
CH	3226.6897	2405.9751	CH	3269.3075	2442.9322
	RS			TS	
ax 2.525 eq 5.0	frequencies/cm^{-1} H_3	D_3		frequencies/cm^{-1} H_3	D_3
Tx+Ry	-47.6971	-42.1351	RCF (Tz)	-229.9506	-215.6082
Rz	82.7151	58.7057	Tx	121.1285	74.0473
Ty+Rx	181.9849	158.1928	Rz	166.7450	110.5002
Tz	532.1784	503.8934	Ty	104.6781	152.0281
Tx+Ry	1117.4639	830.6365	Ry	529.5164	374.5692
Ty+Rx	1133.9003	851.3387	Rx	569.4978	402.8505
scis	1417.4147	1032.9897	opla	1283.2946	968.159
umbr	1430.1430	1052.6533	def	1394.5243	1024.358
rock	1453.2633	1054.1011	def	1403.1651	1031.052
CH	3098.8832	2209.7533	CH (a1)	3086.8517	2183.5724
CH	3239.5245	2412.6081	CH	3302.8863	2466.3494
CH	3244.4179	2417.8638	CH	3303.7752	2467.5706
ax 3.0 eq 3.0					
Tx+Ry	113.0411	100.577	RCF (Tz)	-150.0277	-138.412
Rz	211.3297	185.3978	Tx	220.4983	202.4054
Ty+Rx	299.6085	212.45	Rz	229.3212	210.4789
Tz	488.1557	463.5179	Ty	400.4281	283.2542
Tx+Ry	1128.767	833.106	Ry	411.2725	290.9299
Ty+Rx	1142.668	851.102	Rx	426.7522	301.8753
umbr	1447.191	1052.669	opla	1449.4655	1111.367
scis	1458.938	1072.224	def	1459.7311	1063.883
rock	1480.423	1072.294	def	1462.5842	1066.009
CH	3042.648	2168.508	CH (a1)	2926.1982	2069.93
CH	3189.309	2377.591	CH	3117.5384	2331.725
CH	3191.999	2381.563	CH	3118.8376	2332.843
ax 3.0 eq 3.5					
Tx+Ry	56.6857	50.3059	RCF (Tz)	-149.3962	-137.7529
Rz	192.0591	162.2054	Tx	98.0977	89.7536
Ty+Rx	238.3238	175.0533	Rz	120.7896	110.5016
Tz	509.2439	482.7416	Ty	334.6402	287.5684
Tx+Ry	1129.3118	835.4123	Ry	406.5246	236.7184
Ty+Rx	1144.3734	854.4173	Rx	431.8464	305.4788
umbr	1437.2360	1043.7121	opla	1407.2704	1079.6064

scis	1446.1551	1063.1919	def	1429.8933	1046.1917
rock	1466.8624	1066.6837	def	1432.8328	1048.3846
CH	3057.1380	2179.4569	CH (a1)	2967.3828	2099.0628
CH	3201.9848	2386.4398	CH	3177.5437	2375.1677
CH	3206.0871	2391.2478	CH	3178.3200	2375.9384

	RS			TS	
ax 3.0	frequencies/cm^{-1}			frequencies/cm^{-1}	
eq 3.75	H_3	D_3		H_3	D_3
Tx+Ry	18.5521	16.4416	RCF (Tz)	-156.3286	-144.1103
Rz	182.7040	139.9150	Tx	74.0233	67.6658
Ty+Rx	206.2119	166.7873	Rz	102.2405	93.4449
Tz	522.5221	494.9659	Ty	381.5677	269.9145
Tx+Ry	1129.2334	836.7772	Ry	282.3152	199.7069
Ty+Rx	1144.6849	856.1675	Rx	407.4229	288.2022
umbr	1433.8135	1040.5201	opla	1390.4200	1039.1364
scis	1442.6049	1060.2227	def	1418.4749	1041.4823
rock	1462.7564	1065.3174	def	1421.6246	1066.9363
CH	3066.9933	2186.7319	CH (a1)	2991.7074	2116.2694
CH	3210.5158	2392.2938	CH	3204.4856	2394.4725
CH	3215.1288	2397.4100	CH	3204.9040	2395.0365
ax 3.0					
eq 4.0					
Tx+Ry	-29.5333	-26.1450	RCF (Tz)	-162.3005	-149.5912
Rz	161.7890	116.4075	Tx	65.0902	59.4657
Ty+Rx	189.9689	163.3307	Rz	97.7352	89.2733
Tz	530.7933	502.5168	Ty	356.6918	252.3185
Tx+Ry	1127.7739	836.7990	Ry	234.0498	165.5675
Ty+Rx	1143.3861	856.4613	Rx	385.3527	272.5902
umbr	1430.7695	1038.2195	opla	1377.4614	1057.1604
scis	1439.6824	1058.0315	def	1410.3841	1034.0900
rock	1459.7114	1063.7452	def	1413.5444	1036.4688
CH	3076.4257	2193.6690	CH (a1)	3011.9203	2130.5676
CH	3218.6675	2397.8349	CH	3225.6615	2409.6301
CH	3223.4694	2403.0738	CH	3226.2970	2410.3268
ax 3.0					
eq 5.0					
Tx+Ry	-57.1843	-50.5135	RCF (Tz)	-173.2207	-159.6610
Rz	77.5428	55.0739	Tx	76.5003	69.8267
Ty+Rx	182.6341	158.6643	Rz	106.7900	98.5213
Tz	553.8311	523.3721	Ty	308.727	218.3894
Tx+Ry	1126.5285	838.3122	Ry	112.0628	78.4132
Ty+Rx	1142.6575	858.6469	Rx	342.9821	242.6181
scis	1426.4660	1034.6280	opla	1351.3306	1037.0728
umbr	1435.9968	1054.6935	def	1396.8317	1025.6470
rock	1455.3521	1062.8091	def	1400.2767	1028.2669
CH	3093.7226	2206.6365	CH (a1)	3051.0740	2158.2641
CH	3231.8631	2406.6354	CH	3265.3188	2437.9060
CH	3236.9008	2412.0050	CH	3265.9275	2438.6132

	RS			TS	
ax 4.0	frequencies/cm^{-1}			frequencies/cm^{-1}	
eq 3.0	H_3	D_3		H_3	D_3
Tx+Ry	56.1764	49.7432	RCF (Tz)	-40.8352	-37.4531
Rz	176.6826	128.0887	Tx	204.8559	188.0856
Ty+Rx	198.8693	169.6856	Rz	206.5931	189.6774
Tz	506.791	479.6813	Ty	303.1059	214.411
Tx+Ry	1103.459	817.8449	Ry	309.3926	218.8577
Ty+Rx	1118.82	837.3777	Rx	440.4634	311.5748
umbr	1421.159	1036.969	opla	1461.5988	1064.238
scis	1434.503	1055.279	def	1462.4345	1064.887
rock	1456.442	1056.555	def	1486.3346	1146.342
CH	3075.307	2192.773	CH (a1)	2862.667	2024.99
CH	3217.602	2397.062	CH	3048.0084	2281.403
CH	3221.878	2401.94	CH	3048.9391	2282.049
ax 4.0					
eq 3.5					
Tx+Ry	10.6158	9.3915	RCF (Tz)	18.5875	17.0126
Rz	145.0720	103.5479	Tx	58.6342	53.6574
Ty+Rx	190.8045	165.2167	Rz	67.2817	61.5708
Tz	530.5017	501.7686	Ty	304.9136	215.6906
Tx+Ry	1117.9294	830.0270	Ry	313.0799	221.4660
Ty+Rx	1133.7876	849.9829	Rx	348.9532	246.8420
umbr	1426.6538	1036.1697	opla	1429.9780	1045.7394
scis	1435.9707	1055.7553	def	1430.6965	1046.2581
rock	1456.2241	1061.1117	def	1449.9410	1120.5959
CH	3082.0298	2197.9418	CH (a1)	2918.1008	2064.2018
CH	3222.4512	2400.2760	CH	3128.6179	2339.2861
CH	3227.2120	2405.4609	CH	3129.0910	2339.6590
ax 4.0					
eq 3.75					
Tx+Ry	-30.1901	-26.6954	Tz	-38.3297	-35.0557
Rz	125.3713	89.2427	Tx	-36.6038	-33.4669
Ty+Rx	188.4108	163.5048	Ty	22.9323	20.9665
Tz	539.4340	510.0401	Ry	263.1687	186.1630
Tx+Ry	1121.8195	833.6224	Rx	274.6387	194.2735
Ty+Rx	1137.7149	853.6416	Rz	292.6428	207.0093
umbr	1427.4509	1035.7056	def	1418.2497	1038.5818
scis	1436.6074	1055.3614	def	1419.0804	1039.1864
rock	1455.9846	1062.6674	opla	1432.4023	1107.8655
CH	3085.9183	2200.8818	CH (a1)	2945.2991	2083.4412
CH	3225.4423	2402.2463	CH	3162.3780	2363.4173
CH	3230.3597	2407.5326	CH	3162.7395	2363.7560

	RS			TS	
ax 4.0	frequencies/cm^{-1}			frequencies/cm^{-1}	
eq 4.0	H_3	D_3		H_3	D_3
Tx+Ry	-42.7680	-37.8048	Tz	-62.0896	-55.6595
Rz	109.0205	77.5220	Tx	-172.8773	-158.0895
Ty+Rx	186.6114	162.0584	Ty	-23.0810	-21.0899
Tz	546.1753	516.2268	Ry	67.8660	48.9543
Tx+Ry	1124.0867	835.7564	Rx	236.2460	167.1153
Ty+Rx	1140.1976	855.9616	Rz	240.2921	169.9776
umbr	1427.7536	1035.3803	def	1410.1687	1033.1945
scis	1436.9817	1055.0702	def	1410.9288	1034.2134
rock	1455.7843	1063.6682	opla	1414.3844	1094.4915
CH	3088.1327	2202.5673	CH (a1)	2967.2840	2098.9932
CH	3227.0778	2403.3524	CH	3189.8144	2382.2812
CH	3232.1032	2408.7080	CH	3190.6256	2383.7173
ax 4.0					
eq 5.0					
Tx+Ry	-59.0830	-52.1740	Tz	-74.8606	-68.3901
Rz	48.8434	34.6753	Tx	-363.9759	-333.5281
Ty+Rx	183.5609	159.4891	Ty	33.0725	30.1893
Tz	563.1326	531.6782	Ry	-482.6661	-341.4282
Tx+Ry	1128.2141	840.2121	Rx	133.4394	94.3922
Ty+Rx	1144.6359	860.7512	Rz	104.2114	73.7208
umbr	1427.6389	1034.3007	def	1397.9269	1024.2138
scis	1437.6090	1054.0828	def	1398.8543	1027.1220
rock	1454.9451	1065.6720	opla	1350.7739	1045.9094
CH	3095.2747	2208.0268	CH (a1)	3027.0297	2141.2556
CH	3231.8008	2406.3573	CH	3247.2396	2420.2043
CH	3236.9248	2411.7729	CH	3249.5812	2425.8979
ax 5.0					
eq 3.0					
Tx+Ry	40.5458	35.8055	RCF (Tz)	55.5292	50.904
Rz	80.3359	57.0588	Tx	201.4767	184.9897
Ty+Rx	185.2088	160.8165	Rz	201.7707	185.2559
Tz	541.2767	511.514	Ty	305.9454	216.4192
Tx+Ry	1113.54	828.4603	Ry	306.4788	216.7968
Ty+Rx	1129.319	848.6383	Rx	449.8928	318.2447
umbr	1420.856	1032.844	opla	1460.7613	1063.359
scis	1430.863	1052.5	def	1461.139	1063.668
rock	1451.58	1057.251	def	1492.3759	1151.593
CH	3095.932	2208.143	CH (a1)	2845.0652	2012.538
CH	3234.217	2408.35	CH	3030.5704	2268.843
CH	3239.216	2413.675	CH	3031.0575	2269.181

	RS			TS	
ax 5.0	frequencies/cm^{-1}			frequencies/cm^{-1}	
eq 3.5	H_3	D_3		H_3	D_3
Tx+Ry	-48.5388	-42.8636	RCF (Tz)	76.8928	70.3275
Rz	65.9065	46.7942	Tx	49.7871	45.5605
Ty+Rx	184.7494	160.4859	Rz	52.3511	47.9075
Tz	551.2888	520.7505	Ty	309.8132	219.1552
Tx+Ry	1120.3874	833.9387	Ry	309.0524	218.6170
Ty+Rx	1136.3339	854.1898	Rx	354.0820	250.4700
umbr	1424.3550	1033.5022	opla	1428.6352	1044.6872
scis	1433.7286	1053.2327	def	1428.8551	1044.8326
rock	1453.1309	1061.2366	def	1453.3713	1124.0592
CH	3095.9675	2208.3497	CH (a1)	2904.4632	2054.5548
CH	3233.4859	2407.7121	CH	3118.7069	2332.0686
CH	3238.5614	2413.0904	CH	3119.0033	2332.2939
ax 5.0					
eq 3.75					
Tx+Ry	-52.2687	-46.1538	RCF (Tz)	65.1333	59.5217
Rz	56.5400	40.1381	Tx	-38.7298	-35.4088
Ty+Rx	184.3548	160.1610	Rz	-36.4802	-33.3524
Tz	555.5968	524.7184	Ty	269.2998	190.4969
Tx+Ry	1122.7714	835.9373	Ry	268.7096	190.0794
Ty+Rx	1138.9731	856.3526	Rx	295.7866	209.2331
umbr	1425.3533	1033.6522	opla	1416.8728	1037.5832
scis	1434.8554	1053.3951	def	1417.2088	1037.8046
rock	1453.5695	1062.6556	def	1431.1060	1107.7746
CH	3096.2679	2208.6417	CH (a1)	2934.2992	2075.6601
CH	3233.3551	2407.5490	CH	3156.5976	2359.2010
CH	3238.4640	2412.9513	CH	3156.7069	2359.3199
ax 5.0					
eq 4.0					
Tx+Ry	-53.8217	-47.5231	RCF (Tz)	47.3882	43.2835
Rz	41.7028	29.5988	Tx	-47.9780	-43.8376
Ty+Rx	184.0705	159.9417	Rz	-45.5398	-41.6102
Tz	557.4117	526.4037	Ty	225.4115	159.4512
Tx+Ry	1123.9051	836.8782	Ry	225.2978	159.3717
Ty+Rx	1140.1652	857.3462	Rx	240.8722	170.3878
umbr	1425.7296	1033.6378	opla	1410.1357	1033.8058
scis	1435.2634	1053.3681	def	1410.3609	1033.9495
rock	1453.6255	1063.2392	def	1387.1652	1074.3014
CH	3097.0535	2209.2090	CH (a1)	2971.5829	2102.0338
CH	3233.9925	2408.0008	CH	3196.2709	2387.6157
CH	3239.1339	2413.4190	CH	3196.4497	2387.7775

	RS			TS	
ax 5.0	frequencies/cm^{-1}			frequencies/cm^{-1}	
eq 5.0	H_3	D_3		H_3	D_3
Tx+Ry	-61.9867	-54.7169	RCF (Tz)	-22.3555	-20.4033
Rz	-17.8114	-12.6403	Tx	-24.2204	-22.1093
Ty+Rx	183.1332	159.1018	Rz	-20.8263	-19.0112
Tz	568.7087	536.6694	Ty	111.8716	79.1355
Tx+Ry	1128.4138	840.8688	Ry	109.4719	77.4402
Ty+Rx	1144.8377	861.4611	Rx	102.9832	72.8482
umbr	1427.0125	1033.6704	opla	1396.8856	1025.7794
scis	1437.3782	1066.0015	def	1397.1356	1025.9391
rock	1454.3005	1053.4970	def	1338.1682	1037.1350
CH	3097.3612	2209.6727	CH (a1)	3026.8209	2141.1079
CH	3232.9812	2407.0509	CH	3248.3587	2424.8229
CH	3238.1646	2412.5048	CH	3248.4467	2424.9291

A vast quantity of data is presented in Table 21, however it is particularly important in order to explain the trends observed in the kinetic isotope effects from Tables 19, 20, and associated Figures.

Each of the sets of frequencies corresponds to a methyl cation from a particular pose; a cutoff level 4 equivalent to the structure described in the heading of the frequencies. Therefore, for "*ax 5.0, eq 5.0*", the frequencies correspond for the cutoff level 4 structure (methyl cation) from the original r_{ax}=5.0Å, r_{eq}=5.0Å cage system.

Table 19 showed the errors in *ln*(KIE) for each of the structures and each of the subset Hessians associated with those geometries. It was noted that as r_{ax} was increased, the errors in *ln*(KIE) for the level 4 cutoffs increased also. For continuity, inspection of solely the r_{eq}=3.0Å for values of r_{ax} increasing from 2.04Å through to 5.0Å show that in the TS, the reaction coordinate frequency – which incidentally is the translation in the *z* vector – diminishes in magnitude from 260*i* cm^{-1} at r_{ax}=2.04Å through to 55cm^{-1} at r_{ax}=5Å. It is for values above r_{ax}=3.0Å for which the RCF significantly drops in magnitude, to 41*i* cm^{-1} at r_{ax}=4.0Å, a frequency magnitude which is below the threshold considered relevant in some implementations of isotope effect codes, and would therefore not contribute to the computed IPFRs or IEs.[17] Granted, the contribution of a magnitude corresponding to a modular value of 40 cm^{-1} is comparatively little where the majority of RCFs for values of r_{ax} below 4.0Å exhibit magnitudes of above 100 wavenumbers; it is just this small value which contributes significantly to the size of the errors in *ln*(KIE) for large values of r_{ax}.

As described earlier in this chapter, it is certainly reasonable to expect the subset Hessians for systems originating from "exploded" S_N2 structures, to lead to relatively erroneous results, as compared to more traditional, more condensed geometries where the environmental effect is strong enough to be significantly reflected in the subset Hessian itself, without the explicit influence of the elements relating to the coupled environmental interactions.Indeed, it is also reasonable to consider the exploded structures as S_N1 equivalents, with the isotope effect effectively changing from a KIE to an EIE with the nature of the reaction. As previously described, it is not unusual for an enzyme-catalysed S_N2 reaction to exhibit donor-acceptor distances of approximately 2-3Å, however beyond this the strength of the interactions significantly break down proportionally with distance, leading to an effective reaction conversion from bimolecular to unimolecular substitution in the TS.

Table 22. B3LYP/aug-cc-pVDZ optimised bond lengths (Å) and angles (degrees) for RS and TS

		rO_{eq}				
rO_{ax}		3.0	3.5	3.75	4.0	5.0
2.04	$(OC)_{RS}$	1.583	1.563	1.568	1.563	1.553
	$(CH)_{RS}$	1.0861	1.0896	1.0901	1.0896	1.0885
	$(OCH)_{RS}$	102.2	103.4	103.3	103.4	104.0
	$(CH)_{TS}$	1.0850	1.0882	1.0875	1.0867	1.0850
2.525	$(OC)_{RS}$	1.569	1.563	1.557	1.551	1.540
	$(CH)_{RS}$	1.0929	1.0942	1.0936	1.0929	1.0917
	$(OCH)_{RS}$	103.8	104.2	104.4	104.6	104.9
	$(CH)_{TS}$	1.0964	1.0969	1.0956	1.0944	1.0921
3.0	$(OC)_{RS}$	1.565	1.554	1.547	1.543	1.532
	$(CH)_{RS}$	1.0951	1.0945	1.0938	1.0932	1.0922
	$(OCH)_{RS}$	104.5	104.6	104.7	104.8	105.0
	$(CH)_{TS}$	1.1030	1.1021	1.1004	1.0989	1.0962
4.0	$(OC)_{RS}$	1.553	1.542	1.538	1.535	1.528
	$(CH)_{RS}$	1.0934	1.0930	1.0928	1.0926	1.0922
	$(OCH)_{RS}$	105.1	105.1	105.1	105.1	105.2
	$(CH)_{TS}$	1.1087	1.1065	1.1043	1.1022	1.0974
5.0	$(OC)_{RS}$	1.537	1.532	1.531	1.530	1.525
	$(CH)_{RS}$	1.0921	1.0921	1.0921	1.0921	1.0921
	$(OCH)_{RS}$	105.2	105.2	105.2	105.2	105.2
	$(CH)_{TS}$	1.1100	1.1072	1.1046	1.1013	1.0972

The strength of the C-H bonds varies within each of the cage geometries. The variation in bond strength and geometrical parameters with solvent effect has been a recurring theme throughout this thesis. Indeed, it is one of the factors responsible for the magnitude and direction of a computed isotope effect. Therefore, when applying a cutoff procedure to obtain a subset Hessian, the effect of the environment on the bonds within the system must be inherently considered.

Table 22 shows computed CH bond lengths in the methyl cation for the range of constrained cage structures considered in this work. Consider the trends in C-O and C-H bonds for the structures with r_{ax}=2.04Å. As equatorial C-H···O interactions increase in magnitude (decreasing r_{eq}), the C-H bonds become stronger in accordance with the conclusions from earlier chapters, and the C-O bond in the RS becomes proportionally weaker. In comparison, as r_{ax} increases, changes in r_{eq} begin to have far less influence than in the more axially condensed systems; C-H bond lengths for the methyl cation RS at r_{ax}=5.0Å do not change for any of the values of r_{eq}.

Conversely, it is the C-H bond lengths in the TS which appear to be more affected with increasing values of r_{ax}. At short r_{ax} distances, there is a fluctuation of approximately 0.003Å in the C-H bond length in the methyl cation cage TS structures, however as r_{ax} increases, this variation increases to approximately 0.012Å between values of r_{eq}, *i.e.* a *quasi* four-fold increase in bond length changes. This variation can be attributed to a number of factors. The primary cause is the relatively small effect of the axial substituents at large values of r_{ax}, shown by the breakdown of the RCFs and errors in subset IPFRs in Tables 19 and 21. This suggests that as the influence of the factors in the direction of the reaction coordinate diminish, so do the perpendicular effects increase in influence.

The impact of this affects the application of subset Hessians in two ways. Firstly, as stated previously, it is vital that for a reaction, the effects in the reaction coordinate be maintained; failure of these to be reflected in the subset Hessians result in the large errors observed for cutoff level 4 in the exploded structures in Table 19. Secondly, it is particularly important to appreciate the intrinsic relation between both the RC parallel and perpendicular interactions. As early postulates stated that isotope effects were cumulative, which was proven to be incorrect in certain cases (see Chapter 4), it has been demonstrated in this chapter, that these are indeed not independent. Certainly, each effect arises from the influence of the individual contributions, but is entirely related to the effects in other planes than that which is studied.

This reinforces the conclusions from Chapter 5, which suggest that future calculations and interpretations of IEs from experiment should consider the effects of the entire environment especially for inherently inhomogeneous systems such as enzyme active sites, but also provides important recommendations for the application of subset Hessians – which are inherently valid for a vast number of situations – but which break down for specific cases, and require judicious considerations of the atoms to be retained, in order to reflect the original system.

6.3.2 The Superheavy Cage approach: effect of the cutoff procedure

In order to complete the investigation of the cutoff treatment for the methyl cation water cluster, and its effects on the IEs computed from the cutoff Hessians, the hydrogens of the water molecules were substituted with atoms of atomic mass 999 a.m.u. A disadvantage of the traditional cage approach involved the influence of the spurious water modes, which affected the calculated isotope effects. It had been considered that the water librational modes were responsible for the anomalies detected in the carbon isotope effects in particular, and therefore the cutoff procedure was repeated for the new “Superheavy” system. As described in Chapter 5, this is particularly appropriate when considering the methyl cation within a massive protein environment, the “Superheavy” hydrogens effectively mimicking this. The simplicity of the procedure is remarkable; changing the mass of the hydrogen atoms in UJISO and LIPFR from 1 to 999 a.m.u. has the effect of stabilising the methyl cation relative to its environment.

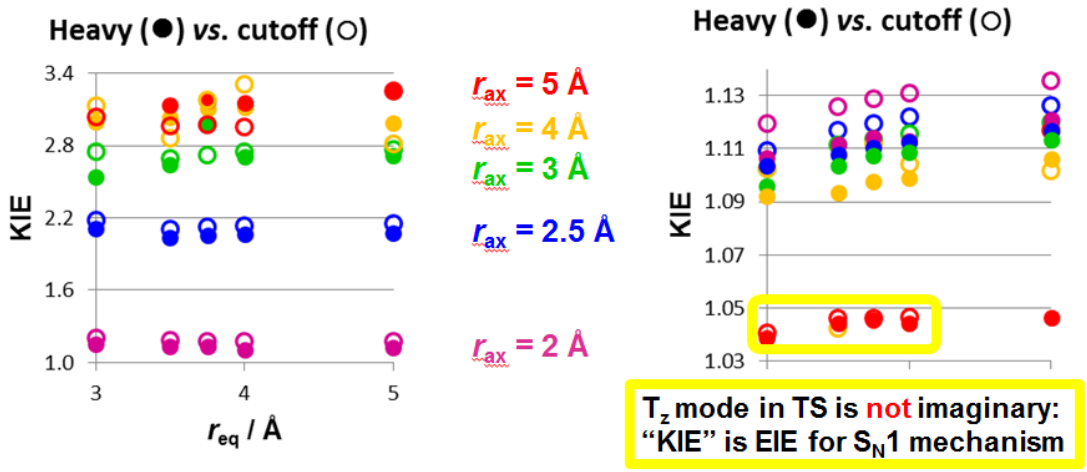


Figure 19. 2° α-D_3 KIEs and ^{14}C 1° KIEs for the identity reaction of methyl transfer between two states of protonated methanol, with substitution of water hydrogen atoms by a superheavy cage.

Very little difference is made to the 2° α-D_3 KIEs, yet there is a drastic improvement in the ^{14}C methyl cutoff (level 4) KIEs. The values from calculations with the full Hessian now being much closer to the cutoff level 4 KIEs. As previously shown in earlier chapters, the level 4 cutoffs tended to produce the most significant errors, described as *ln*(KIE) in Table 20. These were interpreted as arising from truncation of the transition vector, due to the Hessian no longer containing sufficient information in order to reproduce the effect of the environment of the methyl cation, when a cutoff such as level 4 was applied.

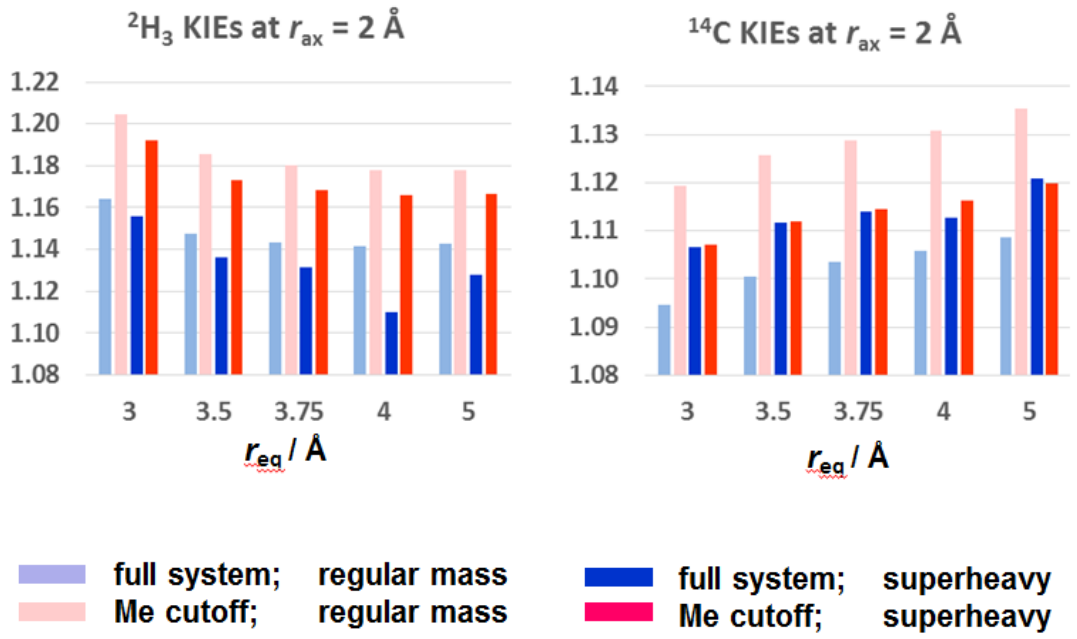


Figure 20. Comparison of cutoff and full Hessian procedures and the application of the superheavy cage approach to the reliability of computed Hessians and isotope effects.

However, it is noteworthy that the ^{14}C KIEs, most significantly interpretable as those more intrinsically linked to the effects associated with the reaction coordinate, are markedly improved within the superheavy approach, as compared to the traditional methyl cutoff. Of course, within the m_H=1 a.m.u. system, the structure and geometry are more of a model, whereas with a change to the superheavy, m_H=999 a.m.u. system the mass effect of a protein or enzyme is replicated by the effective tethering of the otherwise loose hydrogen atoms on the axial and equatorial water molecules.

It is a remarkable similarity that is observed between the superheavy cutoff levels 1 and 4, as compared to the previous definition of cutoff level 1, and level 4, which in the m_H=1 a.m.u. system leads to the marked errors described above. Indeed, whereas the m_H=1 a.m.u. full and methyl cutoffs showed discrepancies of nearly twice the normal magnitude of KIE for the level 4 cutoff, the differences in KIE certainly for r_{eq}<4.0Å are in the region of approximately 3.d.p.

A movement is observed of approximately 1% towards more normal isotope effects, with the superheavy environment effectively decoupling the motions of the methyl

transfer, yet having little effect on the KIEs. There is a nearly 4% increase in the normal direction of the methyl cutoff isotope effects when compared to the full treatment for the 2° α-D_3 KIEs, with the equatorial interactions between 4 Å and 3 Å being approximately 3% more normal for T_3 substitution.[229] In the context of the recently published COMT results for methyl cation transfer by Klinman *et al.* [92], it is believed that this percentage difference could contribute to the KIE differences between the mutant enzymes and the hydrogen bonding differences, as described in additional detail within Chapter 5.

The ^{14}C superheavy cage data shows an increase in the normal magnitude of the KIEs upon moving to the system where m_H= 999 a.m.u., where the small effect can be likened to encapsulation within a protein-like environmental mass *i.e.* an active site. There are approximately 2.5% more normal KIEs for the regular mass system yet little error for the system with superheavy mass.

6.6 Summary

The use of the cutoff procedure has been considered in a number of instances, both with KIEs and EIEs from a simple cage approach to Hessians output from a more supramolecular-approximate structure. Through methodical use of different implementations of the cutoff rules, it has been shown that previously postulated "molecular" methods from the mid-20th century should not be used for subsets of atoms which are significantly coupled to their environments. Issues revolving around misinterpretation of isotope effect data and the effects it has had on the revival of the "*compression hypothesis*" have continued to be discussed in this chapter, building upon the introduction in Chapter 1, and subsequent investigations.

It has been shown that charged species show significant sensitivity to the dielectric/solvent environment, which likely has important implications on the interpretations of experimental KIEs in enzyme-catalysed reactions. The EIE for transfer of the methyl cation from vacuum to water shows good agreement between Hessians with one solvation shell and those containing solely components relating to the methyl cation; 2° α-D_3 KIEs for the transfer of the methyl cation within a water cage being normal with magnitudes varying from small for a highly constrained cluster, to large for the exploded S_N1-like transition structure. This axial distance dominating, the equatorial distances do not cause significant changes in the KIEs due to their relatively small magnitudes.

It has been found that even as small a change as 1 Å in the C – O interactions could result in an α-T_3 KIE becoming more normal by approximately 3% with a 1% inverse effect on ^{14}C. By applying the novel superheavy cage approach, it has been demonstrated that decoupling the methyl cation vibrations from the environment does not change the 2° α-D_3 cutoff KIEs significantly, but markedly improves the 1° ^{14}C cutoff KIEs to the point of 3 decimal place discrepancies between full and cutoff treatments.

It is timely to discuss the importance of the cutoff data and superheavy cutoff data presented in this chapter. Previous instances of the cutoff approximation as applied to systems of any size had relied on the original Bigeleisen formulation for calculation of isotope effects.[227] As described at the beginning of this chapter, this is no longer appropriate for highly environmentally coupled systems such as those under consideration here, for a number of reasons.

In particular, the early idea that isotope effects were effectively additive in nature has been proven to not only be incorrect in some cases; the fact that independent interactions cumulate to cause very subtle changes in isotope effects indicates that the inverse to this rule is in fact true.

Indeed, as has been shown by the investigations of r_{eq} and r_{ax} variation together and independently, changes in each parameter affects the isotope effect arising for the KIE for bimolecular substitution, and in particular, affects this KIE differently for the different combinations of each perpendicular distance.

The subject of models is often discussed as a primarily useful theoretical incentive, whereas moving onto more realistic systems provides more reasonable and lasting conclusions. Of course this is true to a certain extent, however without understanding the behaviour of the small, intrinsically important parts of very large, complex and inhomogeneous systems, it is difficult to establish categorical and reasonable interpretations of the effects observed for these large structures.

Hence, the importance of model system calculations as have been the subject of this thesis thus far. Of course employing a PCM as in Chapter 3 is but an average approximation of a large scale effect, namely the solvent. It is however a useful technique, as was proven with PCM-derived EIEs comparable to QM/MM simulations and reflecting similar reactions in solvent.

Therefore, how has this application of the cutoff procedure to a larger, "model supramolecular system" advanced the understanding of not only the approximation itself, but also the behaviour of the methyl cation in a solvent environment? The data presented within the tables and figures of this chapter, describe the development of the research into the different levels of cutoff applied to a Hessian, resulting in smaller and smaller subsets of the said matrix.

In the traditional, m_H=1 a.m.u. system, with the different combinations of r_{ax} and r_{eq}, the error in the KIEs (described as error in *ln*(KIE) in the tables, as a proportional quantity to the free energy for isotopic substitution) increases with increasing level of cutoff. This is unsurprising, as truncating the full Hessian for a system is guaranteed to have an effect on the output isotope effects, however it is necessary to properly gauge this effect, and the factors responsible for the increase in KIE error with increasing level of cutoff.

The most severe errors were observed for the cutoff level 4 systems, or those where only the rows and columns corresponding to the atoms within the methyl cation, were retained. This was particularly clear for the systems with large values of $r_{ax,}$ where the interactions in the reaction coordinate were weaker than at smaller ranges and therefore a Hessian truncation would eliminate these further from the influences on the methyl cation. Indeed, errors are as large as ~106% at r_{ax}=2.04Å, increasing to figures in the range of 1000% for level 4 also at r_{ax}>3.0Å. Inspection of the reaction coordinate modes confirm that a breakdown in magnitude and direction at large distances results in the increase in error in *ln*(KIE), in particular due to the magnitude and direction influence described in the tables of Chapter 6, where the T_z has a significant contribution to both.

Having acknowledged that the cutoff procedure had very different effects on the cage structures for each cutoff level, transferring the previous investigation into a more supramolecular approximation led to the superheavy description. Where the mass of hydrogen atoms on the cage water molecules is changed from 1 a.m.u. to 999 a.m.u., the behaviour of the system changes. This is unsurprising, however there is a marked effect on the reliability of the cutoff procedure, within the superheavy approximation. Where previously m_H=1 a.m.u., the error in *ln*(KIE) could reach as high as 1000% for the methyl cation cutoff (level 4); repeated calculations under the superheavy approach showed little error when r_{eq}<4.0Å, with larger distances exhibiting errors in KIE of less than 5%.

The importance of this is twofold. Firstly, previous conclusions suggested that a methyl cation cutoff is simply too severe to provide sufficient environmental information to reproduce full system behaviour,[79, 224] whereas the superheavy approach suggests that this is not the case. Indeed, with such small errors in KIE exhibited for an otherwise drastic application of the cutoff procedures, the KIEs obtained are very reasonable.

It is noteworthy that the effect of increasing the mass of the hydrogen atoms on the water molecules to 999 a.m.u. from 1 a.m.u. corresponds to effectively tethering these atoms to a *pseudo* protein environment, reducing/eliminating the number of spurious frequencies associated with hydrogen atom motion within the cage. It is therefore likely that these were responsible for some of the error associated with the level 4 cutoff systems in the m_H=1 a.m.u. calculations; the only difference with the superheavy approach being the increased mass of the hydrogen atoms on the water molecules.

Indeed, the importance of the choice of cutoff theory has been highlighted. A translation-rotation method would effectively project out the librational modes and, should these be isotopically sensitive, lead to errors on application of the cutoff procedure. Conversely, it has been shown that application of the so-called *all frequencies* method can lead to less error in cutoff application, as all degrees of freedom are considered, and small imaginary vibrational modes are treated as normal vibrations. It is important to note the coupling of a system with its environment; highly environmentally-coupled systems will be less appropriate for application of translation-rotation cutoff procedures due to the difficulty in separating the motions of the isotopic system from its environment. Also notable is the description of the reaction coordinate frequency, the accurate description of which is well known to affect the reliability of calculated isotope effects. However, it has been noted that the nature of the reaction coordinate frequency differs with distance, insofar as magnitude decreases with the increasing donor acceptor separation. This has a consequence on the effect of the cutoff procedure. As the distance between donor-acceptor atoms increases, and therefore the reaction coordinate extends, the reaction coordinate frequency decreases in magnitude. This decrease in magnitude has a direct effect on the environmental coupling. A decreased magnitude of RCF has been found to lead to greater sensitivity of isotope effect to cutoff level. Indeed, this is shown in the results obtained herein, where an error of 12% in ln(KIE) is reported for the least extreme level of cutoff and smallest donor-

acceptor distance, compared to over 1000% error in the most extended reaction coordinate. The cutoff procedure may therefore be used when the description of the reaction coordinate remains unaltered, and the interactions with the reacting system are strong enough to be observed in the cutoff Hessian, without being removed entirely in the corresponding rows and columns of forces.

In order to develop this series of model calculations further towards the supramolecular level, QM/MM calculations were carried out on the COMT enzyme structures, which are described in Chapter 7. These calculations are equivalent to the experimental and computational work carried out by other groups, also considering the nature of catalysis of the COMT enzyme.

7. QM/MM Hessians: Computationally obtained isotope effects for enzymic reactions

7.1 Introduction

Described in Chapter 2, QM/MM techniques form an important advance in computational chemistry and biology. By allowing large systems with hundreds to thousands of atoms to be minimised and characterised by different methods, QM/MM techniques have vastly increased the range of structures that can be explored by computational methods, testament going to the 2013 Nobel Prize to Warshel, Levitt and Karplus for the development of these methods.[230]

There has been a focus on model systems in previous chapters. After all, it is important to understand the behaviour of individual parts of a system before considering them in inherently complex and inhomogeneous surroundings. Indeed, the behaviour of the methyl cation in vacuum and explicit solvent has been treated in this work, with the factors responsible for the isotope effect, and environmental influences being described and highlighted. Therefore, in order to develop these ideas further, QM/MM calculations were carried out on the SAM structure in water as the RS, and COMT active site as the TS.

This chapter will describe the results obtained from a number of QM/MM simulations with various electronic structure methods carried out by collaborators at Universitat Jaume I in Castellon, Spain, and processed with the isotope effects and post processing being carried out in Bath. These results are compared to those previously obtained by Klinman and coworkers and re-evaluating these in the context of the compression hypothesis. Additionally, the ideas of modern cutoff implementations for supramolecular systems, described in Chapter 6, will be brought forward and applied to QM/MM Hessians here. The effect of conformation will be tested, as numerous reactant conformations can be observed in such systems. The final part of this chapter will describe other interactions which are often not considered in common calculations, in particular dispersion effects. This will lead to the conclusion of the chapter, and develop into a discussion of dispersion effects in inorganic catalytic systems in Chapter 8.

7.2 Conformational averaging of KIEs and IPFRs

The early QM/MM studies on the coupled hydride/hydron transfer reaction catalysed by LDH that showed different KIEs for different TS structures, each connected to a different RS by means of an IRC path, raised the question as to how these values should be averaged.[231] Possibilities might include the following procedures. The individual KIE could be evaluated for each RS/TS pair, connected by an IRC path, and a simple arithmetic mean could be taken over all the individual values, as indeed was originally done.[59] However, each RS/TS pair was associated with a different energy barrier height, so perhaps the contribution of each individual KIE should be Boltzmann-weighted accordingly? But each RS structure has a different energy, so should the relative populations of each at thermal equilibrium be taken into account? And if interconversion of different RS species is rapid as compared with the chemical reaction, then is the overall rate constant not determined solely by consideration of the relative TS energies?

Truhlar's EA-VTST method with multi-dimensional tunnelling has been highly successful in application to KIEs for enzymic reactions involving transfer of hydride, hydrogen or hydron.[226] The fundamental theory is, of course, applicable to any reaction, even those not involving C–H bond making or breaking in a rate-determining TS, and those in which the contribution of tunnelling is small. Averaging over RS configurations occurs, in the first instance, by means of the umbrella sampling during the determination of the classical potential of mean force. Secondly, instantaneous normal modes and frequencies are determined for each configuration frequencies sampled and average frequencies are obtained and used in the evaluation of zero-point energy and the vibrational partition-function contribution to the free energy. The same is done for TS configurations. Subsequently, the re-crossing and tunnelling transmission coefficients computed for a further sample from the 'quasiclassical TS ensemble' are averaged. At each stage it seems that the arithmetic mean is taken.[74, 232, 233]

The approach developed by Ruiz-Pernía and Williams[234] bears similarity to that of Truhlar and Gao, but is simply an extension of the subset Hessian approach to evaluation of IPFRs and KIEs as described in section 7.1. Partly in view of slightly different terminology that is used, it worth outlining here for the sake of clarity.

First, a one-dimensional PMF is computed with respect to a distinguished coordinate, and then at the free-energy minimum (RS) and maximum (TS), extended QM/MM MD trajectories are followed; the TS is subject to a constraint to prevent it collapsing to the RS or product. From the RS and TS trajectories, numerous independent configurations are extracted at regular time intervals; provided that the MD simulation is adequately equilibrated, this should be a sample from a Boltzmann distribution. For each "snapshot", QM/MM geometry optimization is performed to either a local minimum or saddle-point within a frozen environment. The QM/MM Hessian is computed for a subset of selected atoms within the relaxed region and then the IPFR for each isotopic substitution of interest, for either RS or TS, by means of Equations 35 and 36 involving all $3N_s$ or $3N_s - 1$ frequencies (and the imaginary reaction-coordinate frequency if the one-dimensional tunnelling correction is included). Finally, the arithmetic mean is taken of the IPFRs for the individual RS and TS configurations, and the average (Boltzmann-weighted) KIE is obtained as the ratio of the two values (Equation 38).

$$< k / k' > \; = \; < f_{RS} > / < f_{TS} > \qquad (38)$$

This method considers the IPFRs as thermodynamic (or *quasi*thermodynamic) functions of state; the value of the KIE is independent of the actual path taken from the RS *state* to the TS *state*. Alternatively, n_{RS} x n_{TS} individual KIEs may be calculated for each possible combination of n_{RS} RS and n_{TS} TS configurations, and the arithmetic mean taken over these n_{RS} x n_{TS} pairs. This method also assumes that there is an equal probability of each and every individual RS structure going to each and every individual TS structure. This assumption is entirely consistent with conventional TST but could be inappropriate if relaxation of the protein or solvent environment were slow compared with the chemical reaction step. The individual RS or TS configurations may not necessarily all contain the same number of atoms; in this case it would not be meaningful to perform the averaging upon the individual partition functions, but it is meaningful to take the averages of IPFRs each obtained from isotopologous species with the same number of atoms sharing and the same Hessian. When applied to QM/MM KIEs for six isotopic substitutions for the (non-enzymic) reaction of cyanide anion with chloroethane in explicit DMSO solvent, the means and standard deviations differed very slightly between the two methods of averaging, if enough decimal places were inspected) but were indistinguishable if the number of significant figures was restricted to what was justified by the magnitudes of the standard deviations.[234] The "all possible combinations" method was used in a

study of COMT[80] and subsequently by others;[96, 97] it is perhaps less elegant than the quotient of two average IPFRs (Equation 38), but it does allow the distribution of individual KIE values to be inspected. However, when Equation 38 is used, consideration of the distribution about the mean IPFR values should provide a way to assess whether the limited sample of configurations is adequately representative of the population from which it is taken.

In view of the magnitude of fluctuation in the values of IPFRs for individual RS or TS structures usually found, averaging of IPFRs and KIEs in simulations of enzyme reactivity should be an obligatory requirement. It should no longer be considered acceptable to publish results based upon a single RS structure and a single TS structure unless it can be argued convincingly that these are truly representative of the RS and TS ensembles. Otherwise, when comparison is attempted between calculated and experimental KIEs, there is no way of knowing that the computational result is not an outlier from a distribution about a mean with a significantly different value.

7.3 Does transition state theory still work for KIEs?

The usefulness and validity of modern TST in application to enzyme simulations, including KIE calculations, has been argued with clarity and cogency by Truhlar in a recent review.[235] Despite suggestions that TST was no longer adequate as a basis for understanding KIEs (and their temperature dependence) for enzymic reactions involving significant tunnelling of a transferring hydride/hydrogen/hydron,[236] TST with multidimensional tunnelling and ensemble averaging has been shown to be generally fit for purpose.[237] In particular, the role of non-Boltzmann distributions of vibrational energy (promoting vibrations) has been discussed,[114] with the implication that something beyond TST is required for their full description. However, it has been demonstrated that the complex temperature dependences of KIEs for enzyme-catalysed reactions involving significant quantum tunnelling can be reproduced using a TST framework with inclusion of multiple conformations with different reactivity.[225] QM/MM simulations have been employed with EA-VTST/MT to calculate the "enzyme" KIE on the hydride-transfer reaction catalysed by DHFR, which is the ratio of k_{cat} rate constants for the natural "light" enzyme and the "heavy" enzyme, in which ^{12}C, ^{14}N and non-exchangeable ^{1}H have been replaced by ^{13}C, ^{15}N and ^{2}H.[233]

Good agreement between experiment and calculation was obtained with both wild-type and mutant DHFR for this unusual 2° KIE: the small normal values were attributed to the transmission coefficient factor, reflecting differences in environmental coupling, with no effect on the barrier and the QM tunnelling contribution. The influence of protein dynamics was small and could be reproduced accurately by TST.

Recently, it has been reported that common TST-based methods for enzyme simulation lead to TS ensembles that are considerably different from those obtained from a dynamics-based (transition path sampling) method, thereby raising fresh doubt about the validity of the TST approach.[238] On the other hand, the necessity of having a good reaction coordinate in applications of TST to enzymic simulations is recognised, and new methods have been proposed for TS ensemble optimisation.[239]

7.4 QM/MM Simulations of the COMT enzyme

10 ns AM1 simulations of SAM in TIP3P water for the RS, and the COMT active site for the TS were run using the DYNAMO code by Dr. Maite Roca at the Universidad Jaume I in Castellon, Spain. 100 ps snapshots were taken as starting structures from the MD runs, for minimisation for the QM/MM systems. QM atoms were selected within a 7 Angstrom radius of the SAM moiety. The Hessian and coordinate data were then used in Bath as input for the SULISO suite in order to calculate isotope effects. It has been noted in previous chapters that conformation can also contribute to influencing the isotope effect; the case of SAM is an example of this. Indeed, three conformations have been observed in the RS of SAM in water, each resulting in a range of IPFRs output, based on the number of structures taken from the MD runs for minimisation.

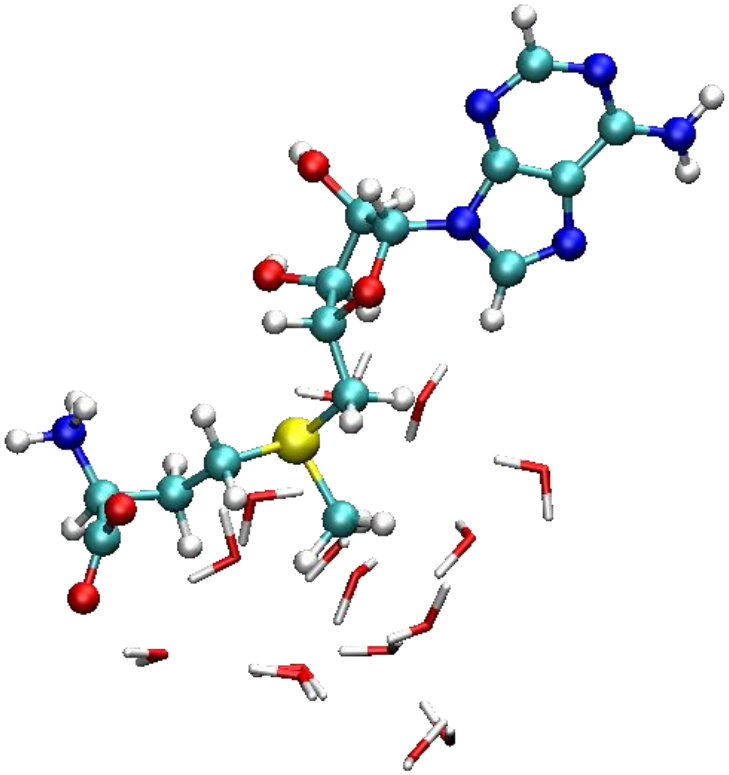


Figure 21. Example conformation of the AM1 RS, SAM in water, in the *extended* conformation. The QM region is shown, where the MM region is described by TIP3P water.

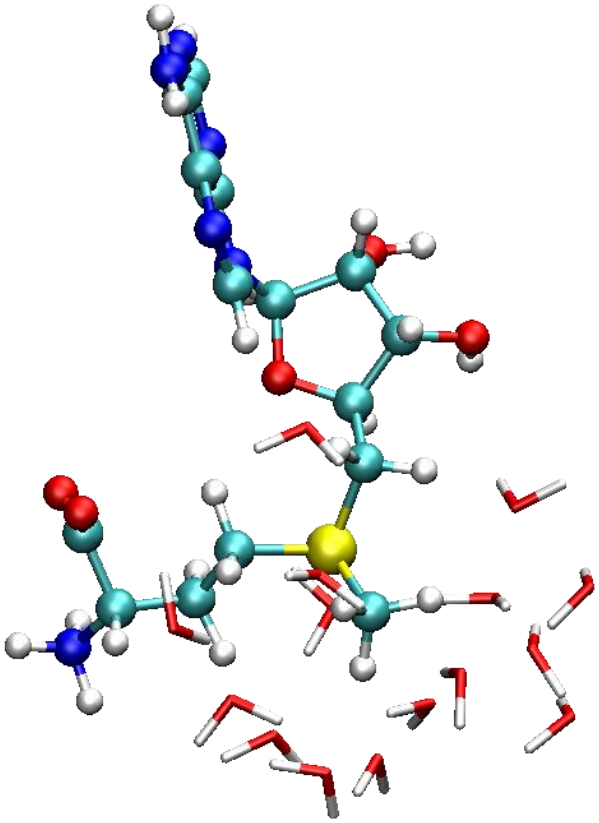


Figure 22. Example conformation of the AM1 RS, SAM in water, in the *partially-folded* conformation. The QM region is shown, where the MM region is described by TIP3P water.

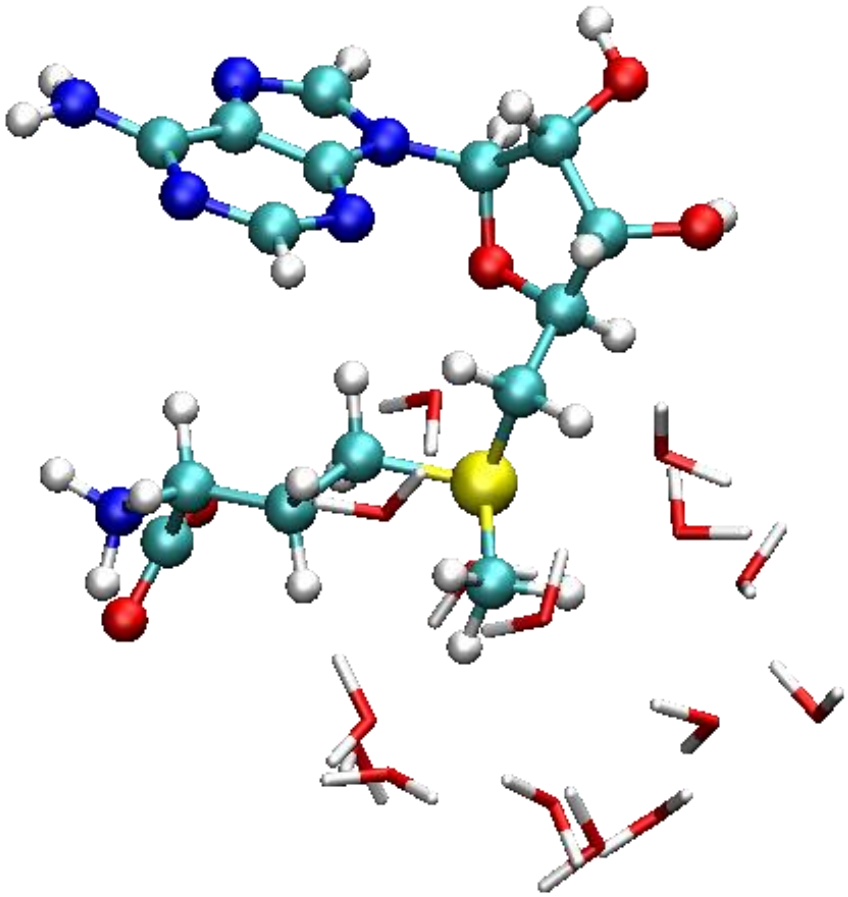


Figure 23. Example conformation of the AM1 RS, SAM in water, in the *folded* conformation. The QM region is shown, where the MM region is described by TIP3P water.

Figures 21 to 23 show representative geometries for each of the three conformations of the SAM structures in the COMT RS used to calculate the IPFRs in the following tables. It can be seen that numerous interactions differ between the conformations, likely being responsible for differences in the calculated properties and IPFRs for the systems.

Table 23. 2^{o} α-D_3 IPFRs and contributing factors for the *extended* conformation.

RS Structure number	IM	EX	ZP	IPFR
1	0.04	1.63	18647.75	1349.84
2	0.04	1.63	17438.55	1261.74
3	0.04	1.76	15602.66	1219.30
4	0.04	1.66	17682.04	1304.35
5	0.04	1.68	17341.90	1290.07
6	0.04	1.58	18439.42	1290.59
7	0.04	1.70	16693.06	1261.76
8	0.04	1.56	20680.08	1429.47
9	0.04	1.57	19189.44	1339.68
10	0.04	1.62	18271.66	1309.07
11	0.04	1.66	17851.40	1317.96
Average	0.04	1.64	17968.32	1305.20

Table 24. 2° α-D_3 IPFRs and contributing factors for the *partially-folded* conformation.

RS Structure number	IM	EX	ZP	IPFR
1	0.04	1.72	15998.64	1221.67
2	0.04	1.66	17796.52	1310.23
3	0.04	1.58	18462.78	1293.00
4	0.04	1.68	17160.79	1278.54
5	0.04	1.66	16872.12	1243.38
6	0.04	1.64	18084.56	1312.19
7	0.04	1.72	16991.39	1295.98
8	0.04	1.66	17324.47	1277.59
9	0.04	1.66	17315.34	1274.04
10	0.04	1.62	18300.43	1314.71
11	0.04	1.63	17693.08	1275.60
12	0.04	1.57	20777.94	1443.13
13	0.04	1.70	16673.12	1257.01
14	0.04	1.70	16313.74	1229.96
15	0.04	1.70	16073.00	1213.11
Average	0.04	1.66	17334.07	1282.68

Table 25. 2^{o} α-D_3 IPFRs and contributing factors for the *folded* conformation.

RS Structure number	IM	EX	ZP	IPFR
1	0.04	1.69	16976.91	1272.14
2	0.04	1.69	17052.36	1275.72
3	0.04	1.67	17906.07	1323.92
4	0.04	1.61	18827.37	1340.80
5	0.04	1.73	16510.80	1266.48
6	0.04	1.64	17616.09	1279.23
7	0.04	1.67	16411.11	1216.00
8	0.04	1.70	16519.90	1242.30
9	0.04	1.66	17864.41	1314.65
10	0.04	1.63	18558.95	1338.62
11	0.04	1.64	18259.97	1326.42
12	0.04	1.64	17456.58	1273.22
13	0.04	1.69	17451.58	1307.42
Average	0.04	1.67	17298.34	1290.53

Tables 23 to 25 feature the IPFRs calculated for the extended, partially folded, and folded conformation of SAM in water, the RS for COMT-catalysed S_N2 methyl transfer for α-D_3 substitution on the methyl cation.

The number of structures representative of each conformation varies due to the gradient tolerance limit in the QM/MM simulations; these post-process structures being those with more reasonable geometries.

As can be expected, the MI factor remains the same for each of the conformations, depending on the magnitude of the mass effect. The excitational factor (EX) appears to on average, increase with increased folding of the SAM structures. The ZPE is, as described in previous chapters, the major contributor the isotope effect, and carries a representative decrease in magnitude on going from extended to partially folded, to folded. At extended conformations, the average ZP factor decreases from $17.97x10^3$, to $17.33x10^3$ at partially-folded, to $17.29x10^3$ for the folded conformation.

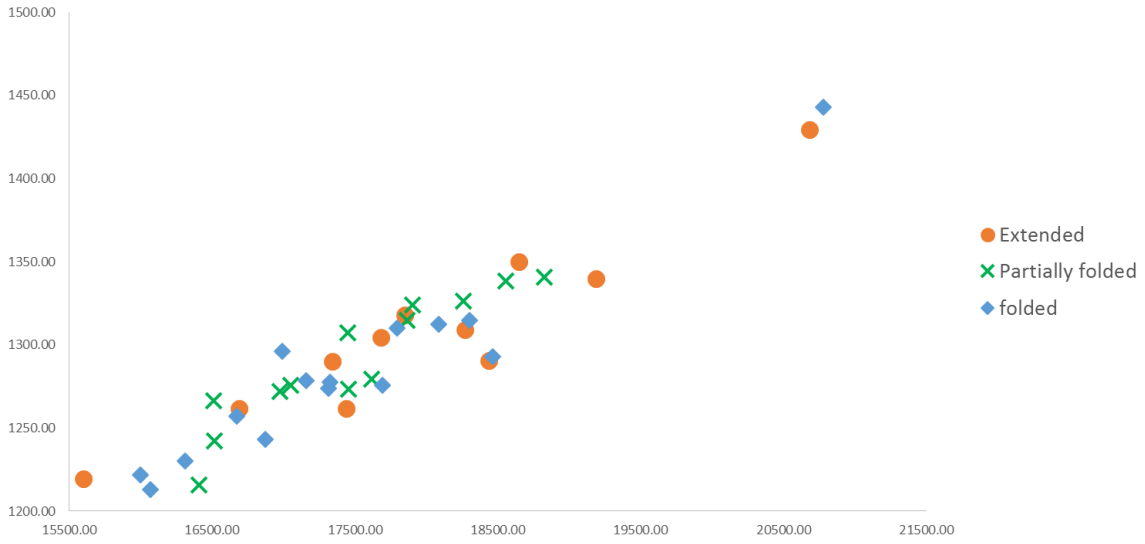


Figure 24. Correlation between the IPFRs and ZP factors, respectively for α-D_3 substitution on the methyl group in SAM with water.

Good linear correlation is obtained from plotting the raw IPFRs against the corresponding ZP factors for each of the SAM in water systems taken as RS structures for the methyl transfer reaction catalysed by COMT. The partially-folded structures follow an appreciable distribution, however removing the outliers for the extended and folded conformations results in improved linear regression for each of the respective system subsets.

The subject of conformations of the flexible moiety in the RS has been the subject of previous publications.[240] There are, of course variations in the magnitudes of the IPFRs for each of the different conformations of SAM in water, however these small changes in the IPFR have little effect on the KIEs obtained.

Table 26. ^{14}C 1^{o} KIEs and standard deviations for extended, partially-folded and folded SAM conformations.

Conformation	KIE	σ
Extended	1.110	0.003700
Partially-Folded	1.108	0.004282
Folded	1.109	0.003390

The changes in KIE for the different conformations are small, and following previous recommendations in the literature regarding isotope effect precision,[241] each of the KIEs effectively equates to 1.11 when considered to three significant figures.

These ensemble-averaged structures exhibit a variety of solvent conformations within the series corresponding to a particular solute conformation. Indeed, the IPFRs reflect the total environmental setting of the structure, with some variation around the ensemble-averaged values in Table 26.

Table 26 also presents the standard deviation in each of the ensemble-averaged values for the SAM in water conformations. With some variation in the values of σ, the KIEs would consistently remain at 1.1 when considered to a single decimal place. It is therefore reasonable to consider that within the current group of structures, the conformation of SAM does not affect the KIE significantly enough to lead to erroneous results depending on the RS.

7.5 Applying the cutoff procedure to QM/MM Structures

In section 7.4, the effect of the COMT methyl transfer RS on the isotope effects produced from QM/MM simulations was considered. Through an ensemble averaging procedure, it was concluded that for the AM1 results presented, the conformation of the RS did not affect the KIE to a significant extent, and should therefore not be considered responsible for any differences in isotope effect direction and magnitude from experiment and higher-level theoretical methods.

The cutoff procedure has been the subject of a significant portion of this thesis, and it is therefore timely to explore its application to systems other than the models considered previously. Not only does this give us information on the validity of the cutoff procedure itself, but also on the methyltransferase reaction, the environmental effects which need to be reproduced in order to continuously obtain reasonable IEs from cutoff Hessians, but also the validity of cutoff when compared to traditional theory, not based on supramolecular systems.

In order to compare the cage model cutoff application with that of a larger more biologically-relevant system, coordinates and Hessians were obtained from QM/MM calculations carried out by collaborator Dr. Maite Roca, in a box of TIP3P water using the M06 functional and 6-31G(d) basis set. Four groups of Hessians were considered; the full system, 3-bond, and 2-bond cutoffs according to Wolfsberg and Stern, and the strict methyl cutoff considered previously.

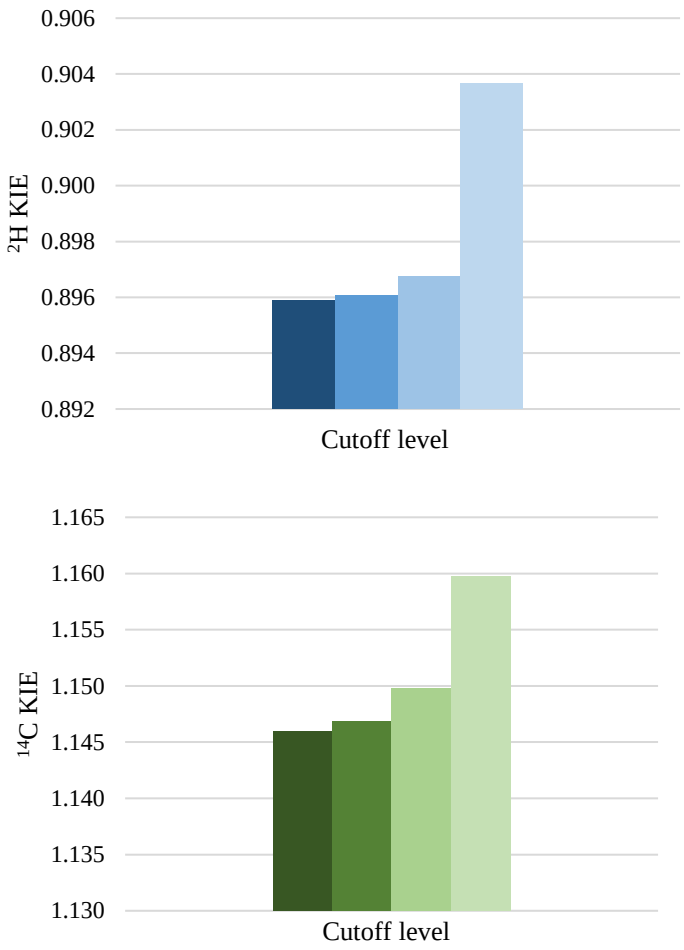


Figure 25. 2° α-D_3 and primary ^{14}C KIEs obtained from the methyl transfer from SAM to catecholate. Darkest colour indicates the full Hessian, lighter the 3-bond cutoff then 2-bond cutoff, and palest the more drastic methyl cutoff.

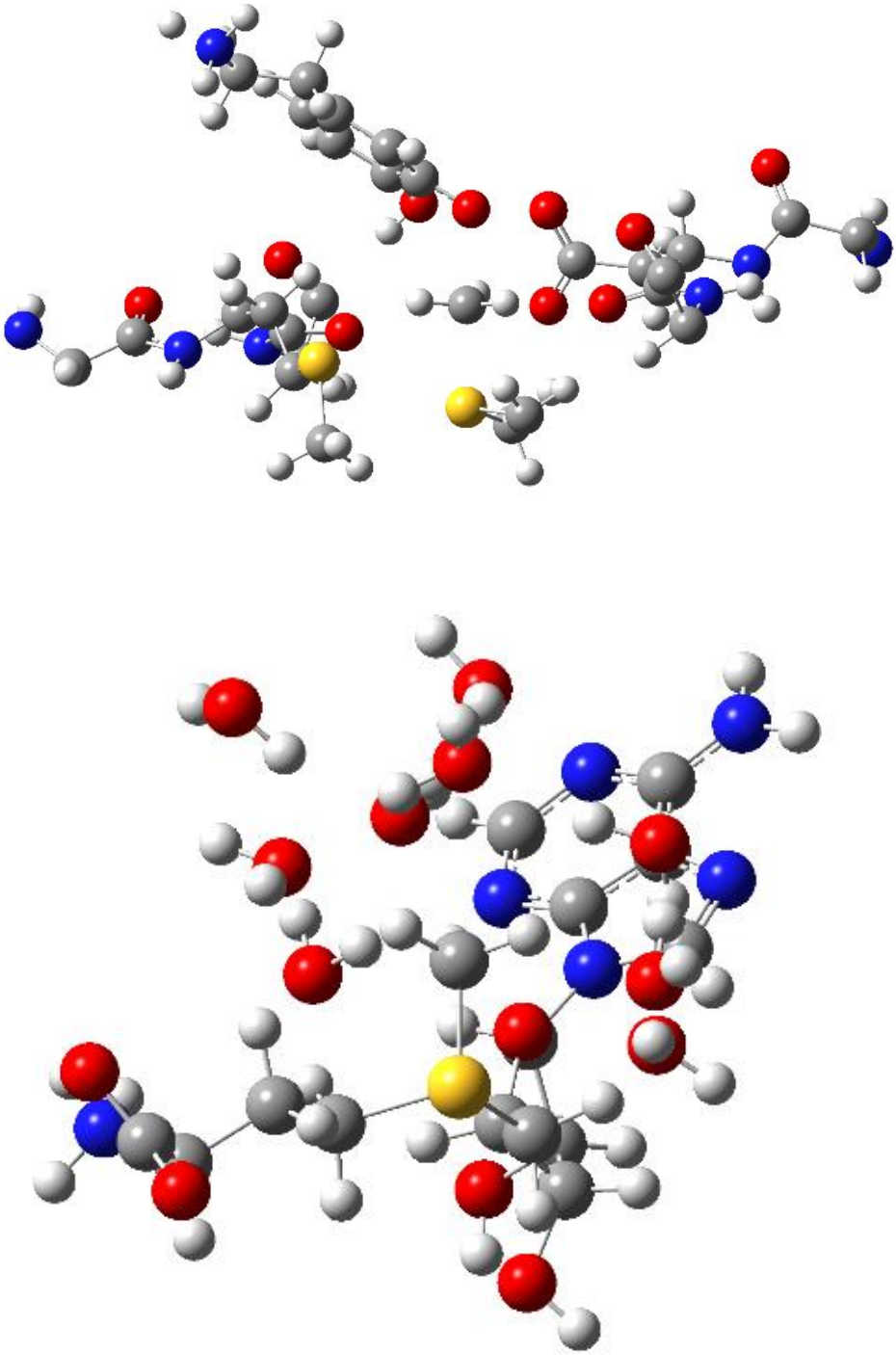


Figure 26. Selected TS and RS poses, respectively from the QM/MM simulation.

It can be seen from data obtained in the cutoff procedure in earlier chapters, that the more severe a cutoff level (such as methyl cation cutoff/ level 4), the less accurate a cutoff IE becomes due to significant Hessian truncation.

Therefore, a similar trend is reproduced here with the QM/MM Hessians; the more significant methyl cutoff predicts the isotope effect to the least accuracy, whilst still predicting the correct direction and relative magnitude. Going to a 3-bond cutoff as was suggested by Wolfsberg and Stern has little effect on the KIE compared to the full system. The 2-bond cutoff performs relatively well, however the more drastic cutoff to the methyl cation (0 bond cutoff for deuterium and 1 bond cutoff for ^{14}C) raises the KIE for both isotopologues, toward a more normal result.

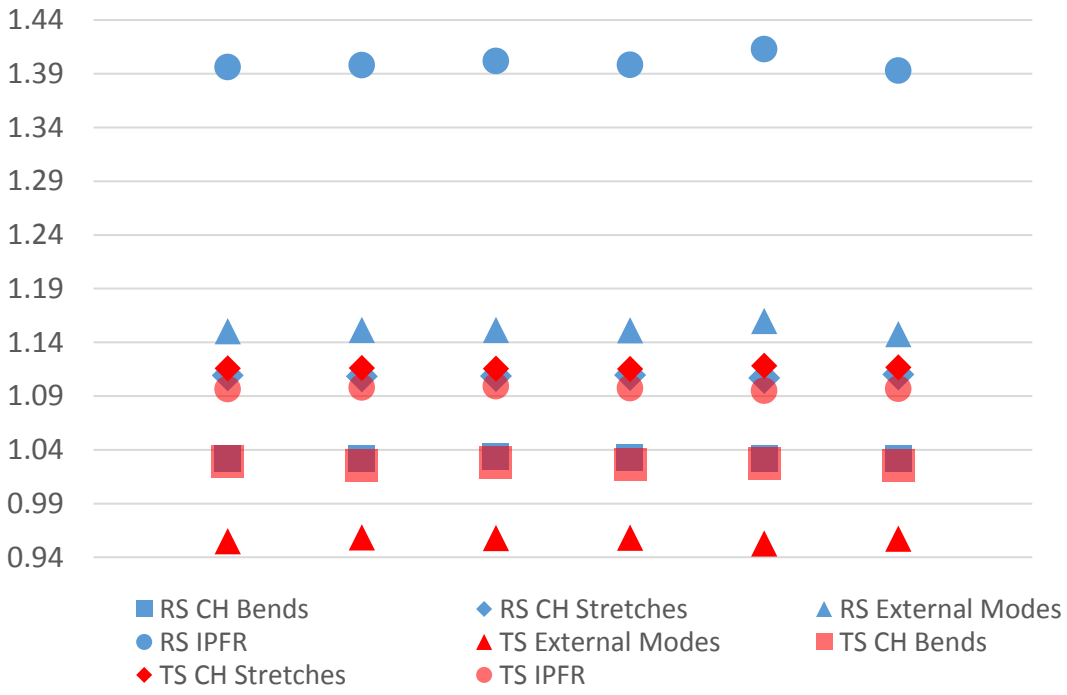


Figure 27. Individual components to the ^{14}C primary IPFR for RS (blue) and TS (red) structures output from a QM/MM simulation of methyl transfer within the COMT active site, where each point on the *x*-axis represents a different structure.

Figure 27 contains the CH stretch, HCH bend, and external mode components to the IPFR for the QM/MM RS and TS structures. The CH stretches and bends are relatively constant in magnitude and trend, however it is the external modes which appear to govern the trend and direction of the IPFR, in particular the translation in the *z*-vector in the TS. As has been seen in the cage structures from earlier chapters, the transition structure frequency is inherently affected by environmental influences and therefore it is intrinsically necessary for its factor to be considered in terms of application of the cutoff procedure. Although it is evident that the complete methyl cutoff is too harsh, it is possible to discern the importance of the reaction coordinate and its correct description in the isotope effects output from our calculations.

Applying the cutoff procedure sequentially to our QM/MM systems, going from methyl cutoff to the full system, a significant improvement in the isotope effects from the cutoff Hessians when applied within a three-bond limit from the isotopically-substituted atom is observed. Although too harsh, the explicit frequency analyses of the methyl cation cutoff structure highlights the importance of correctly describing the reaction coordinated frequency, which is responsible for the trends observed in IPFRs and therefore isotope effects.

7.6. Dispersion effects in simulated enzyme reactions

Dispersion effects have long been a consideration when carrying out electronic structure calculations, but their impact on isotope effects themselves has seldom been considered [242].

London dispersion forces were first postulated as being additive [243]. So for say, three molecules, the forces could be estimated based on the sum of the interactions between Molecule 1 and 2, 1 and 3, and 2 and 3. In recent years, with more conclusive methods testing the basis of dispersion, it has been suggested that the interactions are more complex than simply being additive [7]. This can be simply and primarily explained by the influence of the environment and therefore its contribution to the influence dispersion has on a system [244].

In terms of applicability to organic molecules, it was in 1956 that Pitzer *et al.* commented on dispersion being responsible for the isomerisation energy differences in branched and linear alkanes [245]. For the supramolecular systems discussed here, it has been pointed out that the larger the molecules, the greater the influence of

dispersion interactions in terms of geometry and reactivity, although there is still a noticeable effect on smaller systems. Interestingly, this renewed interest in dispersion has led to some remarkable discoveries; H-bonding, rotational alkane behaviour and especially the repulsive nature of tert-butyl groups have all been reconsidered in a dispersive context [8]. Particularly interesting is the sterically-driven argument of tert-butyl repulsion. When considering the effect of tert-butyl groups, the leading thoughts pertain to their steric bulk and therefore repulsive nature. In their recent discussion of the effects of dispersion, Norrby and coworkers identified the significant impact of dispersive interactions on the in fact attractive nature the tert-butyl group can exhibit [246].

Due to this additional information and the general move towards dispersion-corrected DFT, it seemed appropriate to apply this to our QM/MM simulations, particularly with the H-bonding networks present in the solvated systems. However, considering the nature of the dispersion calculations, it has been pointed out that basis-set superposition error, or BSSE could play a role in causing overstabilisation of complexes. This arises when orbitals from adjacent moieties overlap and effectively appropriate their respective basis functions, resulting in lower energies [247].

BSSE can be a significant cause of error in DFT calculations with small basis sets [248], however the use of the counterpoise correction which accounts for these effects, and incorporating larger basis sets with diffuse functions, generally reduces the BSSE to an unremarkable level [249].

7.7 Dispersion and COMT

It is widely recognised that most DFT methods tend to lack sufficient incorporation of dispersion [250]. This is hardly surprising, given the nature of LDA and GGA functionals. Considering that the energy is obtained based on the local electron density for LDA, and the energy gradient in GGA functionals, mainly long-range effects such as dispersion will inherently not necessarily be parameterised [103].

However, certain corrections and method combinations have been found to provide adequate descriptions of this effect. It is interesting to note that the 6-31G(d) basis set, in combination with B3LYP produces results which have been suggested to *simulate* dispersion [251]. By not actually including any dispersive terms in their parameterisation, the BSSE and dispersive elements effective combine to produce

results akin to those with dispersion corrections. This however is not perfectly reliable, as a cancellation of effects does not constitute an accurate consideration of dispersion. Grimme and coworkers have made suggestions to improve DFT and in particular, B3LYP results in order to correct for this unusual combination of errors.[252]

Indeed, Mulholland and coworkers have pointed out that the B3LYP functional, as well as a number of other functionals, do not include terms for the attractive interaction described as dispersion.[253] Including dispersion corrections can have a profound effect on the energy barriers or enzyme-catalysed reactions, as well as providing results closer to experimental values.[254, 255] Visser and coworkers considered the effects of employing different electronic structure methods, with differing contributions from dispersion, on the barriers and proximity to experimental results of oxygenation processes in enzymes. Their major recommendation involved using a minimum of tripe-zeta basis sets, as well as dispersion corrections with DFT functionals, in order to capture the complex interactions involved in their reactions involving transition metals bound in the active site.[256]

Considering the QM/MM systems earlier in this chapter, reactant structures of SAM, and the COMT-catalysed methyl transfer between SAM and catecholate TS were used.

Table 27. Components to the IPFR from B3LYP/6-31G(d) and B3LYP/6-31G(d),GD3BJ (dispersion included) optimisations. TS indicates data relating to the transition structure of COMT-catalysed methyl transfer, and RS the reactant structure. D_3, T_3, ^{13}C, and ^{14}C refer to trisubstituted deuterium and tritium, carbon-13 and carbon-14, respectively. MI, EX and ZP are the moment of inertia, excitational and zero-point contributions to the IPFR, and QC being the quantum correction for the TS IPFR based on Bell's model for tunnelling.[20]

Structure	Isotope	MI	EX	ZP	(QC)	IPFR
TS (dispersion)	D_3	4.43E-02	1.54E+00	4.42E+04	1.00E+00	3.02E+03
	T_3	7.21E-03	2.18E+00	4.60E+06	1.00E+00	7.21E+04
	^{13}C	2.16E-02	2.26E+00	9.63E+03	1.00E+00	4.70E+02
	^{14}C	7.93E-01	1.07E+00	1.31E+00	1.02E+00	1.09E+00
TS	D_3	4.43E-02	1.54E+00	4.39E+04	1.00E+00	3.00E+03
	T_3	7.21E-03	2.19E+00	4.56E+06	1.00E+00	7.16E+04
	^{13}C	2.16E-02	2.25E+00	9.58E+03	1.00E+00	4.65E+02
	^{14}C	7.93E-01	1.07E+00	1.31E+00	1.02E+00	1.09E+00
RS (dispersion)	D_3	4.43E-02	1.63E+00	3.74E+04		2.70E+03
	T_3	7.21E-03	2.39E+00	3.60E+06		6.19E+04
	^{13}C	2.16E-02	2.90E+00	7.36E+03		4.60E+02
	^{14}C	7.93E-01	1.10E+00	1.43E+00		1.25E+00
RS	D_3	4.43E-01	1.60E+00	3.76E+04		2.68E+03
	T_3	7.21E-03	2.34E+00	3.63E+06		6.13E+04
	^{13}C	2.16E-02	2.89E+00	7.35E+03		4.58E+02
	^{14}C	7.93E-01	1.10E+00	1.43E+00		1.25E+00

Overall, the IPFRs and their components vary little between dispersion-optimised (DO) and dispersion corrected (DC) calculations. It is important to highlight the meaning of the terminology employed. Each optimisation was carried out identically, the only difference being the dispersion correction specified within the optimisation routine for the DO calculations. This may seem intuitive, however a great deal of discussion has taken place on the subject of *when* and more importantly, *how* dispersion corrections should be applied [257-259].

Although inherently not called in the SCF procedure, the dispersion corrections as applied to Gaussian09 have a profound effect in terms of optimisation routines and output geometry.[6] This was shown recently in terms of pi stacking and aromatic systems, where a different structure is obtained when dispersion is employed within the optimisation. Indeed, Matthews and coworkers have considered the effects of including dispersion with a variety of DFT functionals on the structure and interactions of imidazolium ionic liquids. Not only were the optimised structures particularly sensitive to the functionals chosen, but also to their combination with D3 or D2 dispersion corrections, yielding varying pi- and H-bonding interactions.[260] Although it has been seen that dispersion applied as single-point corrections following the initial small basis set calculation can produce relatively reasonable energies, it has been suggested that these are likely due to error cancellations for systems. Conceptually, if a structure would theoretically be affected by dispersion effects and a geometry is deemed incorrect (by not being DO) and output energetics are reasonable, the results are more likely coincidence than from a rational dispersion optimised geometry. This is seen in the Ruthenium catalysis work presented in Chapter 8, where all degrees of freedom vary significantly between DO and DC optimisations.

The difference being observed here however, is the inherent significant H-bonding influences of the structures. With water molecules and a number of both H-donating and accepting fragments; these energetic influences are likely significantly dominating the energy landscape, mediating the effect of dispersion interactions themselves.

Table 28. Gibbs activation energies, $\Delta G^{\ddagger}$, KIEs for DO and DC structures for methyl transfer catalysed by the COMT enzyme, and experimentally derived data.

	$\Delta G^{\ddagger}$ (kcal mol^{-1})	D_3	T_3	^{14}C
DO: B3LYP/6-31g(d) + GD3BJ	12.5	0.895	0.858	1.146
DC: B3LYP/6-31g(d)	12.9	0.893	0.855	1.144
Experimental	18.5[91]	0.83±0.05[207]	0.791[91]	1.063[91]

The isotope effects themselves are testament to the discussion above, with little difference between the DO and DC structures. It is a particularly useful test to have carried out, particularly in the light of the increased scrutiny such methods achieve from reviewers, especially when conjecture such as that relating to the compression hypothesis still remain very much at the forefront of the field. It can be suggested that the methods used thus far offer little difference to calculated isotope effects by incorporating dispersion within the optimisation routine, and with judicious use of scaling factors, it is reasonable to continue with these protocols (as well as our other chosen ESMs including MP2), in order to further develop our understanding of enzyme catalysis. It is however worth noticing the difference in values of $\Delta G^{\ddagger}$ for the B3LYP/6-31g(d) and B3LYP/6-31g(d) + GD3BJ methods. Inclusion of dispersion corrections lowers the barrier for COMT-catalysed methyl transfer. Previous work by Roca *et al.* [261] presented $\Delta G^{\ddagger}$ for the AM1 semiempirical and MP2 post-HF methods as part of QM/MM simulations, of 10.6 and 15.0 kcal mol^{-1}, respectively. As has been discussed previously in this section, the B3LYP method with the 6-31G(d) basis set has particularly recently been shown to benefit from corrections to account for attractive van der Waals interactions and associated errors.[252] It can be suggested that the influence of hydrogen bonding interactions accounted for by the GD3BJ keyword in the optimisation could stabilise the transition structure of COMT-catalysed methyl transfer, thereby reducing the barrier compared to the same calculation with simple energetic corrections.

The issue remains however, that as described earlier, the B3LYP functional has been found to provide results which under certain circumstances approximate the effects of dispersion well, due to pure cancellation of errors.[262] Of course this has been shown to cause little difference within KIEs for COMT, however other structures and systems have been shown to have both geometries and energetics significantly affected by inclusion of dispersion.[254] Future investigations will consider alternative electronic structure methods.

7.8 Summary

In this chapter, the isotope effect, and applications of the cutoff procedure, in a more supramolecular context have been considered. Indeed, the conformations of the inherently flexible RS of SAM in water were computed in order to gauge the effect of the conformational changes in terms of KIEs produced from these AM1 results. It was found that in the context of the current work and the ensemble-averaging procedure, the conformation of the RS structure resulted in a difference in IPFRs, however when taking the quotient of RS and TS IPFRs, this difference was mediated by the TS factor, and resulted in the same directions and magnitude of KIEs to at least two significant figures.

A cutoff analysis of methyl transfer in COMT followed, with RS structures based on SAM in water, and TS structures as a subset of the atoms within the active site of the enzyme, all from M06/6-31G(d) calculations. The conclusions for our model system in Chapter 6 were confirmed in the context of these structures, with accuracy based on the reaction coordinate frequency being responsible for the decline in KIE quality in the most severe cutoff of the methyl cation.

Indeed, using traditional methodologies considering the cutoffs in terms of number of bonds removed from the atom of isotopic substitution, the three bong cutoff suggested by Wolfsberg and Stern in the first part of the 20th century, as being the accuracy limit for cutoff procedures, certainly provides reasonable approximations to the KIEs obtained from the full Hessians of the structures. Following on from this, gradual decreases in the sizes of the Hessian matrices led to degradation of the KIEs obtained, with an *all frequencies* comparison of the methyl cation isotopologue again highlighting the importance of the external modes, in particular translation in the *z*-vector, or the reaction coordinate frequency. It is therefore encouraging that similar conclusions can be drawn upon this work, as that in Chapter 6, showing the readily transferrable concept of the methyl cation cage model into a more supramolecular context.

As a final investigation into factors affecting isotope effects in enzyme-catalysed reactions, the subject of dispersion corrections will be further treated in Chapter 8. This came about from work carried out by Schreiner earlier this decade, suggesting

that traditional interpretations of dispersion could be both lacking in specificity and wholly marginalising smaller-scale interactions. Indeed, we have shown that structurally, dispersion can have an effect in the COMT catalysed methyl transfer to catecholate, affecting the barrier, however causing little change in the isotope effects. This discussion will be expanded in the following chapter where an inorganic catalytic system will be treated and the effect of dispersion considered.

This will be followed by chapters summarising the work presented in this thesis, providing guidelines for the calculation of theoretical isotope effects and interpretation of general isotope effects, and finally an outlook on current avenues in the innovations of isotope effect theory and its applications particularly to biologically-relevant problems. New calculations using the B3LYP functional and incorporating explicit dispersion corrections within the calculation routine, instead of as single point corrections throughout have been presented. This has proven important as cancellation of errors within this electronic structure method shows little effect from explicitly specified dispersive interactions within the routine. However, it is important to highlight that this is effective only within this method and for the current structures. Therefore, Chapter 8 will describe a study on ruthenium-catalysed C-H activation; a catalytic cycle not unlike that which can be conceivably formulated for enzyme catalyses such as COMT. The effect of including dispersion corrections within the catalytic cycle and through a different ESM shows differences in the structure geometries, but also output thermodynamic and kinetic data.

8. Isotope effects in inorganic catalysis

8.1 Introduction

In conversation with colleagues in the organic synthesis department, it was understood that they were looking to supplement one of their projects with computational modelling and isotope effect predictions. Structure optimisations were carried out by C. McMullin, who had previously worked on similar systems.[242, 263-267] Structures were then computed to produce the Hessians and Cartesian files and ran these through our programs in order to obtain the kinetic isotope effects for the reaction.

The aim was to provide a broad prediction of the rate determining step in the catalytic cycle. As is detailed below, the answer was far less evident than expected, with a number of additional considerations which needed to be taken into account. From previous work, the C-H activation was supposedly the RDS [268-280], however incorporating novel ideas from the Kozuch group [281], it was found that additional factors often ignored in previous work needed to be considered in this scenario. Indeed, the main questions of this study centred around the nature of the rate determining step/state and its effect on the observed catalysis. Also, the contribution of each individual step would be of scientific and computational interest due to their contributions to the activity and observed isotope effects. Additionally, and developing upon the work described on dispersion effects in enzyme systems, the nature of dispersion as applied to these inorganic catalytic systems would reveal the impact of this effect on geometry and calculated KIEs.

8.2 New perspectives on isotope effects in catalytic cycles

Kozuch postulated that the concept of rate determining steps, should in fact be reconsidered as *rate determining states* [282]. This is based on the concept that one transition structure and one intermediate can be the major influence in a catalytic cycle. He proposes a definition [281]: *"Rate-determining states are the transition state and intermediate which exert the strongest effect on the over-all rate with a differential change on their Gibbs energies"*. However, what does this mean in the context of traditional catalytic cycles?

The rate-determining states are chemically defined as the TS and intermediate exhibiting the *"highest degree of rate control"*. The degree of rate control was defined as the change in rate with Gibbs energy, and is therefore somewhat an energetic function. This degree of rate control is exhibited with a traditional catalytic cycle by likely a single *state* or two *steps*, as previously described. Within any catalytic cycle, the number of cycles within a given period of time is defined as the turnover frequency (TOF), which can be described within the steady state approximation.

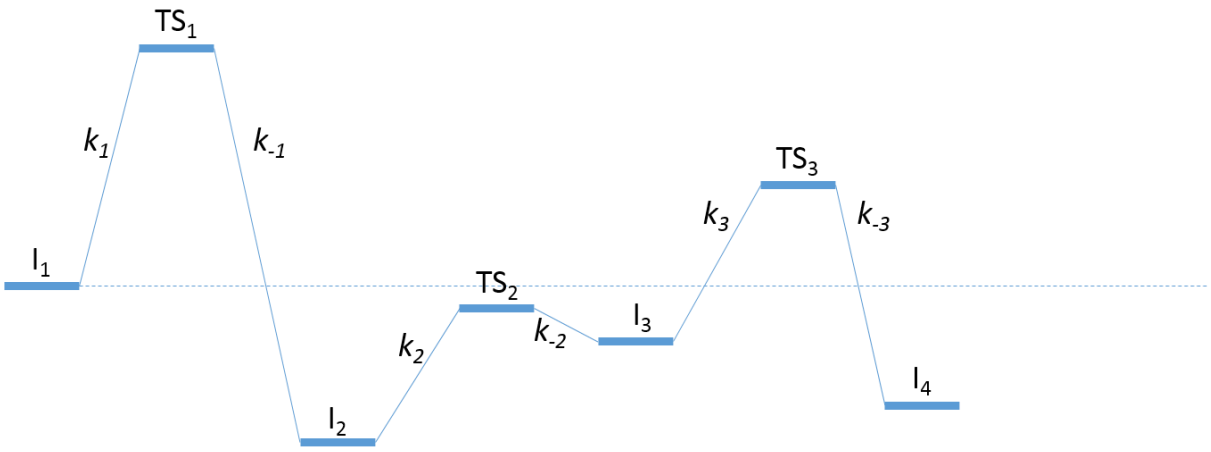


Figure 28. Example three step reaction.

In a three step reaction, as in Figure 28, there would be three rate constants for formation, and three for decomposition of the subsequent state. The formula for this reduces to:

$$TOF = \frac{k_2 k_3}{k_{-2}} = k_3 K_{eq} \tag{39}$$

Where K_{eq} is the equilibrium constant for the formation of TS_2, given by k_2/k_{-2}. As can be seen above, the rate constant for the first step, k_1 is no longer present after cancellation, and therefore does not enter into the TOF determination. However is this an accurate representation of the situation if k_1 were a so-called rate determining rate constant?

Kozuch therefore suggested working in energetic, as opposed to rate constant form, considering more than one cycle in order to gauge the TOF, and RD-state more effectively. This way, the barrier to reaching the products (also reactants for the second cycle) can be gauged, as well as the highest energy TS over multiple cycles.

He also suggested that this methodology, incorporating TS theory, provides more reliable information on catalytic cycles than the single-cycle steady state determination which is used traditionally.

Therefore the effect of incorporating a second cycle was considered, and the largest energy difference between an intermediate and subsequent transition structure (largest effective barrier) in the cycle.

Added consideration is given to the *d*G factor of the exothermic nature of catalysis, therefore bringing the energy of the second cycle down. In practice, this is effectively more conceptually appropriate for the notion of cyclic processes, insofar as the nomenclature of a reactant and product become blurred.

Our isotope effect calculations corroborated the observed isotope effect and through incorporating intrinsic isotope effects and stepwise and conversion corrections, it was possible to justify the results obtained, which are presented and discussed in the following sections.

What follows is an introduction to the reaction through a mechanistic and energetic overview of the cycle, followed by discussion on the isotope effects obtained from that methodology. This is then developed into a dispersion investigation.

8.3 Isotope effects in inorganic catalysis: ruthenium-catalysed C-H activation

8.31 Methodology

BP86[283, 284] calculations were run with Gaussian 09 (Revision D.01).[174] Ru centers were described with the Stuttgart RECPs and associated basis sets[285] and 6-31G** basis sets were used for all other atoms.[286, 287] Initial BP86 optimizations were performed using the 'grid = ultrafine' option, with all stationary points being fully characterized via analytical frequency calculations as either minima (all positive eigenvalues) or transition states (one negative eigenvalue). IRC calculations and subsequent geometry optimizations were used to confirm the minima linked by each transition state. All energies were recomputed with a larger basis set featuring cc-pVTZ on Ru and 6-311++G** on all other atoms. Corrections for the effect of 2-Me-THF (ε = 6.97) solvent were run using the integral equation formalism polarizable continuum model.[128] Single-point dispersion corrections to the BP86 results

employed Grimme's D3 parameter set with Becke-Johnson damping[288] as implemented in Gaussian.

BP86 calculations with the above basis sets were undertaken to establish the mechanism and energetics of the reaction for 3-phenyl-2-oxazolidinone (1a) and methyl acrylate (2b). Previous studies have shown the importance of using corrections for dispersion, solvation and extended basis sets when treating large organometallic reaction systems that involve charged species.[263] A similar approach was adopted in this work, with geometries initially optimized in the gas phase with the BP86 functional and a medium sized basis set (SDD for Ru and 6-31G** on all other atoms). The resultant free energies were then corrected for solvation (2-Me-THF), dispersion (Grimme's D3-BJ parameter set) and an extended basis-set (cc-pVTZ for Ru and 6-311++G** for all other atoms) giving rise to composite free energy differences ΔGMeTHF (1 atm, 25 °C) relative to complex A and separate species 1a and 2b.

8.32 Mechanism and discussion

In order to provide some context to the isotope effect calculations presented in the following section, the mechanisms and energetics inherent in the catalytic cycle are described below. These calculations were carried out collaboratively for publication in *ACS* Catalysis, between synthetic and computational colleagues.[6] Tables and associated figures are extracted from the publication in *ACS Catalysis:* J. A. Leitch, P. B. Wilson, C. L. McMullin, M. F. Mahon, Y. Bhonoah, I. H. Williams and C. G. Frost, ACS Catalysis, 2016, DOI: 10.1021/acscatal.6b01370.[6]

Under the catalytic conditions, acetate from the copper complex ($Cu(OAc)_2{\cdot}H_2O$) breaks up the dimer $[RuCl_2(p\text{-cymene})]_2$ to form the in situ catalytically active intermediate A, [Ru(p-cymene)(OAc)(1a)]+, with the silver compound ($AgSbF_6$) removing chloride anions from the solution. The cationic complex A has the oxazolidinone coordinated through the carbonyl oxygen, one κ^2-acetate and an η^6 para-cymene ligand around the ruthenium center.

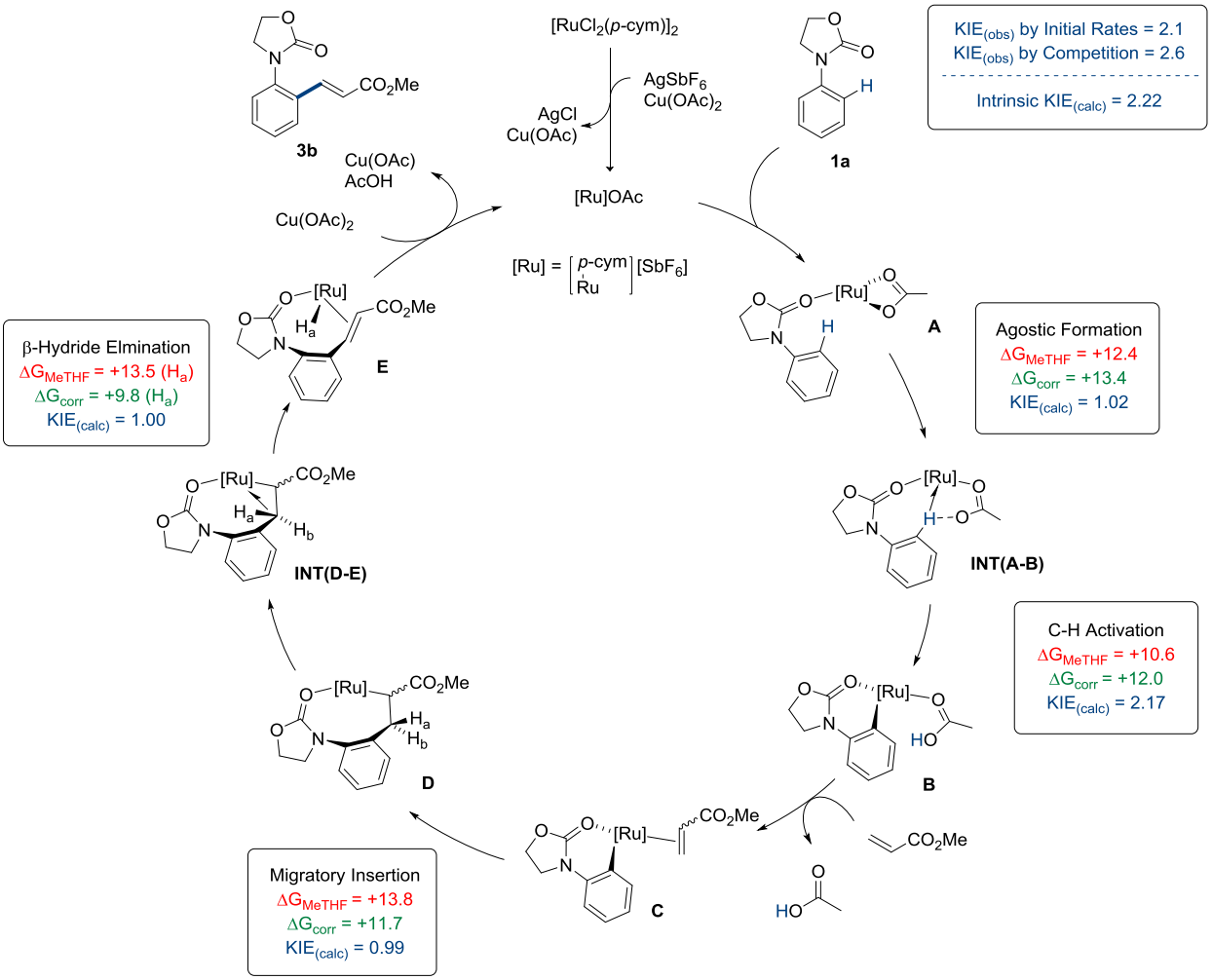


Figure 29. Full Mechanism of Ruthenium-Catalyzed C-H Alkenylation including Kinetic and Energetic (kcal mol^{-1}) BP86 data. Graphic from J. Leitch, P. B. Wilson, *et al.*, *ACS Catalysis* , 2016, *6, pp 5520-5529.*

Concerted metallation-deprotonation (CMD), also known as ambiphilic metal-ligand activation (AMLA), occurs as a two-step reversible process (see Figure 30). The first step, via TS(A-B)1, involves η^2- η^1 displacement of acetate, by the approaching ortho C-H bond of 1a, to form an agostic intermediate INT(A-B), where the pendant oxygen of the acetate is directed towards the ortho H (O···H = 1.686 Å), thereby elongating the C-H bond from 1.091 Å to 1.148 Å. The second step, via TS(A-B)2, involves endergonic C-H bond cleavage to form a six-membered cyclometalate B (+4.0 kcal mol^{-1}). Formation of the agostic intermediate, which involves breaking a strong Ru-O bond, determines the overall C-H activation barrier of 12.4 kcal mol^{-1}.

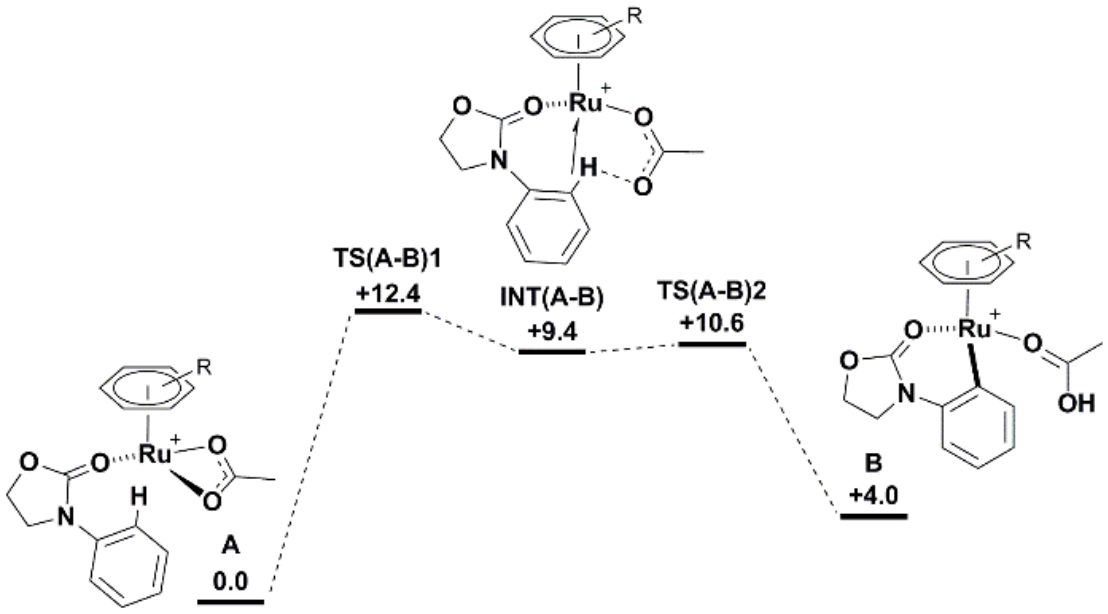


Figure 30: DFT calculated free energies (kcal mol^{-1}) relative to A for the C-H activation of *N*-aryloxazolidinone 1a at $[Ru(OAc)(p\text{-cymene})]^+$ in Me-THF. Graphic from J. Leitch, P. B. Wilson, *et al.*, *ACS Catalysis* , 2016*, 6, pp 5520-5529.*

Ligand substitution of acetic acid in B for methyl acrylate 2b forms four isomers of C, which differ with respect to the orientation of the alkene at the ruthenium center. Despite the pre-1,2-insertion intermediates C11,2 and C21,2 being more stable than the equivalent pre-2,1-insertion intermediates, the free energy barriers for 1,2-insertion (TS(C1-D1)1,2 and TS(C2-D2)1,2) are respectively 20.0 and 19.9 kcal mol^{-1}, approximately 5 kcal mol^{-1} higher than 2,1-insertion (see Figures 31 and 32).

Therefore, 2,1-insertion of 2b into the Ru-C bond is regiosterically favored via TS(C1-D1)2,1 (13.8 kcal mol^{-1}) and TS(C2-D2)2,1 (14.9 kcal mol^{-1}), placing the methyl ester substituent next to the ruthenium in the exergonic eight-membered metallacycles either below (D12,1; -1.7 kcal mol-1) or above (D22,1; -1.3 kcal mol-1) the plane of the ruthenacycle (when looking at the complex from the position of the para-cymene ligand).

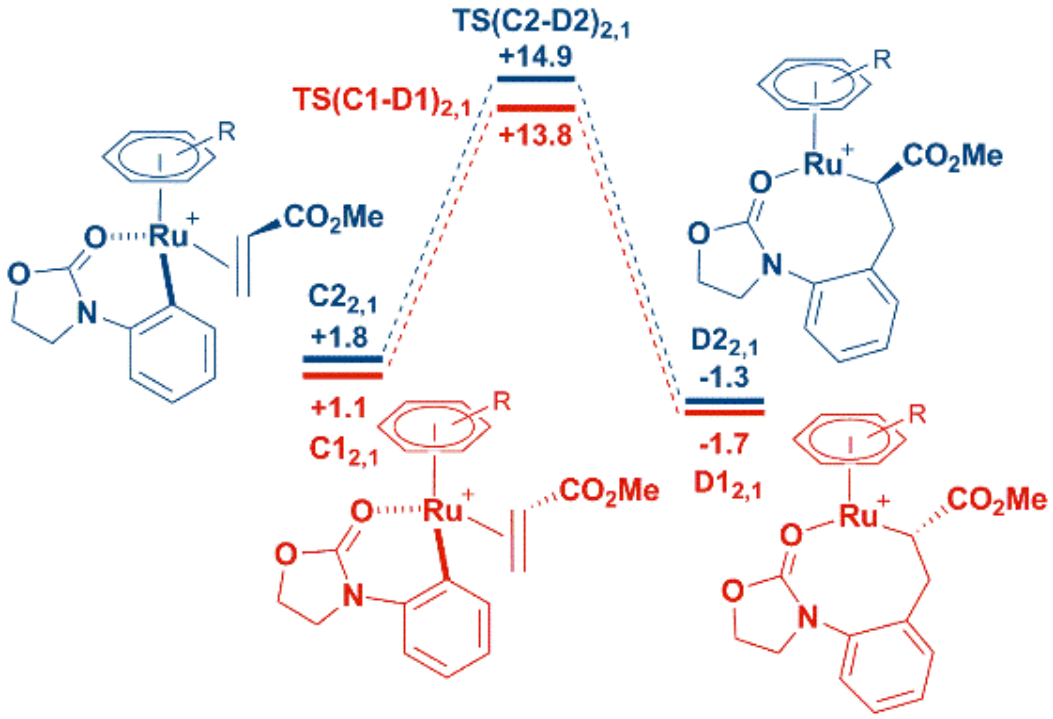


Figure 31: DFT calculated free energies (kcal mol^{-1}) relative to **A** and the free substrates, for the 2,1-insertion of methyl acrylate **2b** into adduct **C**. Graphic from J. Leitch, P. B. Wilson, *et al.*, *ACS Catalysis* , 2016, *6, pp 5520-5529.*

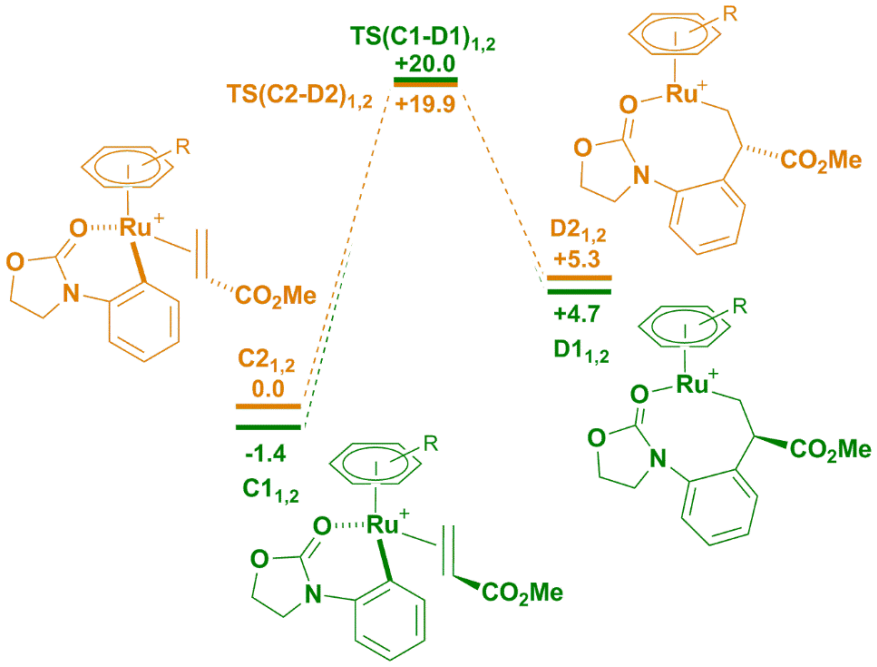


Figure 32. Computed reaction profile (kcal/mol) for the 1,2-insertion of methyl acrylate (**2b**) into adduct (**C**), in Me-THF. In each case free energies are quoted relative to **A** and the free substrates. Graphic from J. Leitch, P. B. Wilson, *et al.*, *ACS Catalysis* , 2016*, 6, pp 5520-5529.*

Unlike in previous functionalization studies with methyl acrylate (by Davies and Macgregor, who used 3-phenyl-pyrazole to form a seven-membered rhodacycle)[242] for this 2,1-insertion no interaction is observed between Ru and the ester substituent in the eight-membered ruthenacycle. In fact, the increased size of the metallacycle restricts its ability to inter-convert between conformers D12,1 and D22,1, as the preference for the boat-chair conformation reduces the flexibility of the metallacycle ring; this contrasts with the behavior that has been reported for similar, yet smaller, equivalent seven-membered intermediates.[264] The stereochemistry of the 1,2-disubstituted alkene product 3b is determined exclusively by which β-hydrogen is transferred to the ruthenium center; from either above (H_a, Figure 33) or below (H_b) the plane of the ruthenacycle.

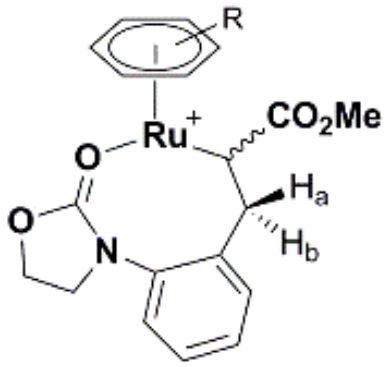


Figure 33: Ruthenacycle $\mathbf{D_{2,1}}$ showing the position of H_a (above) and H_b (below) the plane of the metallacycle. $\mathbf{D1_{2,1}}$ has the CO_2Me substituent *cis* to H_b whilst $\mathbf{D2_{2,1}}$ has the ester group *cis* to H_a

The β-hydrogen transfer process involves two steps: formation of an agostic intermediate followed by C-H cleavage, as the β-H moves to the Ru center. Figure 34 shows the transfer of Ha, which lies above the plane of the ruthenacycle, for both D2,1 isomers. In the case of D12,1, where the transferring hydrogen is trans to the ester group, formation of the agostic interaction between Ha and Ru proceeds via TS(D1-E1)1a (19.1 kcal mol^{-1}) to give intermediate INT(D1-E1)a (+13.1 kcal mol^{-1}), and C-H cleavage occurs via TS(D1-E1)2a (13.9 kcal mol^{-1}) to form the cis product E1cis (+10.4 kcal mol^{-1}). For D22,1, with Ha cis to the ester group, formation of the agostic interaction proceeds via TS(D2-E2)1a (13.5 kcal mol^{-1}), decreasing the Ru···Ha distance from 3.701 to 2.716 Å in INT(D2-E2)a (+5.8 kcal mol^{-1}) before C-H cleavage via TS(D2-E2)2a (7.8 kcal mol^{-1}) and formation of the experimentally observed trans 3b product [Ru(p-cymene)(3b)(H)]+ (E2trans; +4.9 $kcalmol^{-1}$).

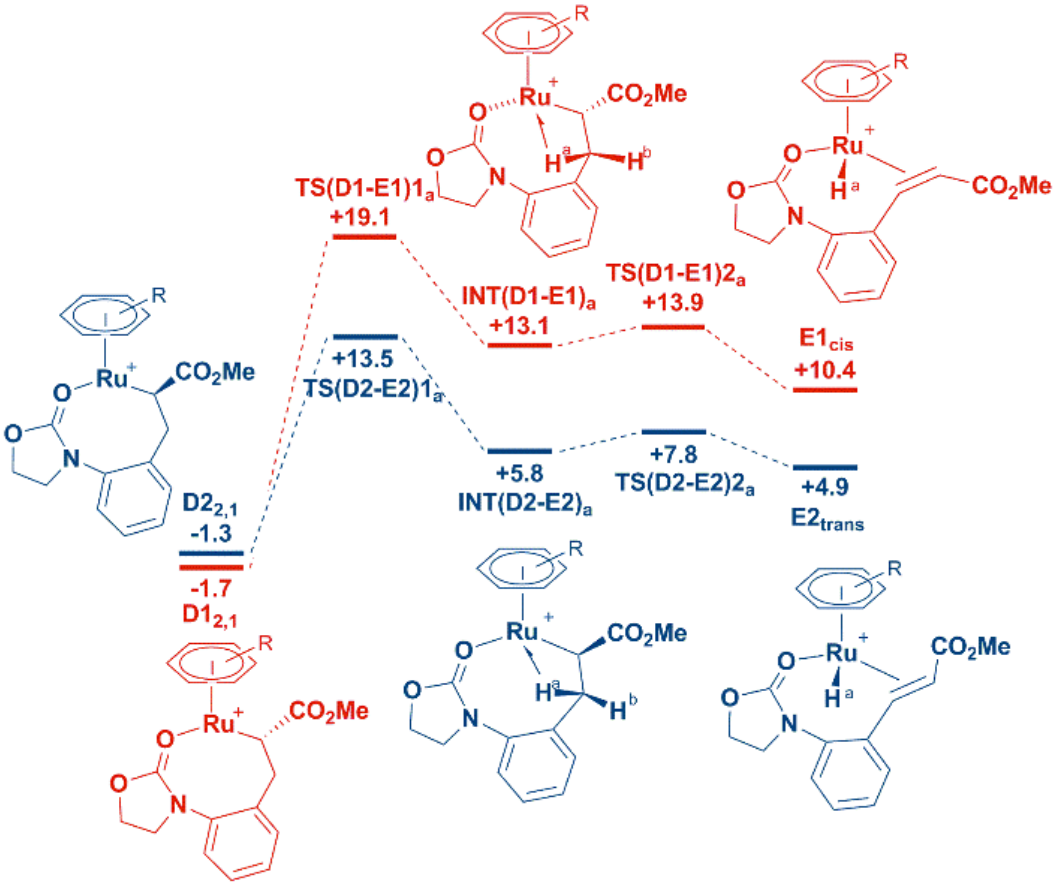


Figure 34: DFT calculated free energies (kcal mol^{-1}) relative to **A** and the free substrates, for the β-H_a transfer from 2,1-insertion ruthenacycles (**$D_{2,1}$**) in Me-THF. Graphic from J. Leitch, P. B. Wilson, *et al.*, *ACS Catalysis* , 2016, *6, pp 5520-5529.*

The process of β-H transfer is more complicated for the "bottom" hydrogen, H_b shown in Figure 35, as the formation of the agostic interaction between H_b and Ru (Ru···H_b decreasing to 2.4 Å) forces dissociation of the oxazolidinone group at the ruthenium center (Ru···O increasing to 3.4 Å). This raises the barriers for the first step from D12,1 via TS(D1-E1)1b (17.5 kcal mol^{-1}) and from D22,1 via TS(D2-E2)1b (31.9 kcal mol^{-1}). The C-H_b bond is elongating by ~ 0.03 Å as the Ru-Hb distance decreases; this is a greater distortion than C-H_a (~ 0.01 Å) when the agostic intermediate is formed for the top β-H transfer. C-H_b cleavage occurs via TS(D1-E1)2b (13.6 kcal mol^{-1}) and TS(D2-E2)2b (20.4 kcal mol^{-1}) for D1 and D2 respectively to give intermediates E1′trans (+9.7 kcal mol^{-1}) and E2′cis (+14.1 kcal mol^{-1}). Reassociation of the oxazolidinone oxygen occurs via TS(E1′-E1)trans (12.8 kcal mol^{-1}) and TS(E2′-

E2)cis (16.8 kcal mol^{-1}) to form the equivalent E1trans (+11.3 kcal mol^{-1}) and E2cis (+14.3 kcal mol^{-1}) complexes, which are less stable than the preceding E′ complexes.

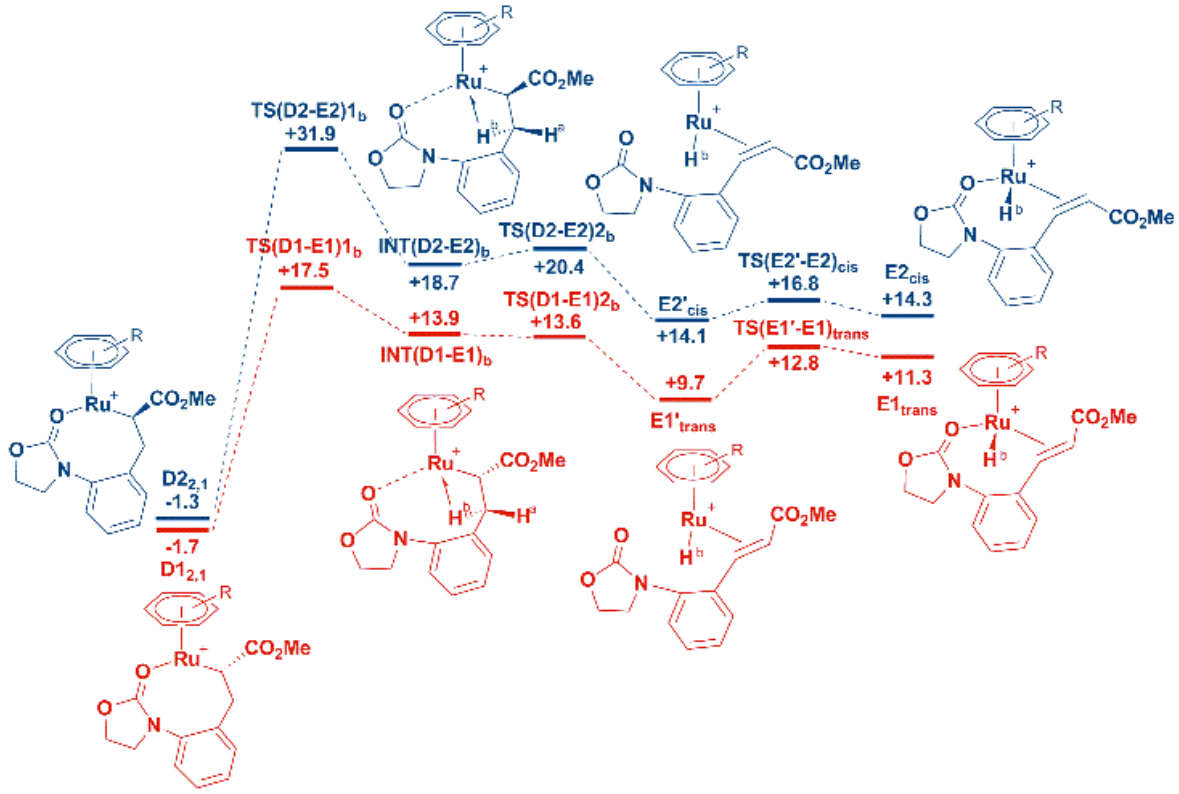


Figure 35: DFT calculated free energies (kcal mol^{-1}) relative to **A** and the free substrates, for the β-H_b transfer from 2,1-insertion ruthenacycles (**$D_{2,1}$**) in Me-THF. Graphic from J. Leitch, P. B. Wilson, *et al.*, *ACS Catalysis* , 2016, *6, pp 5520-5529.*

Based on the assumption that the energy difference of 0.4 kcal mol^{-1} between the two 2,1-insertion ruthenacycle complexes (D12,1 and D22,1) is small enough that both species are populated during the catalytic cycle, it is comforting to note that both ruthenacycle isomers preferentially form the trans 1,2-disubstituted alkene product over the cis stereoisomer. This agrees with experiment that only trans 3b is observed, and is due to H transfer of different β-hydrogens (H_a or H_b). The free-energy difference between TS(D1-E1)1a and TS(D2-E2)1a suggests a preference of about 104 for formation of the trans product over the cis isomer.

The absence of a di-alkenylated product (3bdi) formed during the reaction was investigated, with the subsequent C-H activation of 3b at the remaining ortho C-H

position modelled. The barrier for this activation was 15.4 kcal mol^{-1} (see Table 29), 3 kcal mol^{-1} higher than activation of the initial C-H site in substrate 1b (TS(A-B)1 = +12.4 kcal mol^{-1}), hence 3b does not undergo a second C-H activation. The unusual isomeric preference 1ar / 1aq substrates were likewise studied. Here, the major and minor C-H activation pathways were modelled for 1ar (see Table 30) and unsurprisingly the barrier for the major isomer pathway was lower in free energy, by 0.9 kcal mol^{-1}. This small difference, and similar energies for 1arA and 1arA′, the major and minor isomers of [Ru(p-cymene)(OAc)(1ar)]+ respectively, (ΔG = 0.4 kcal mol^{-1}) agree with the experimental observation of both isomers, showing a slight energetic preference for the formation of the major isomer.

Energy contributions for Tables 29 to 31

ΔE_{BS1} SCF energy computed with the BP86 functional with BS1

ΔH_{BS1} Enthalpy at 0 K with BS1

ΔG_{BS1} Free energy at 298.15 K and 1 atm with BS1

$\Delta G_{BS1/MeTHF}$ Free energy corrected for 2-MeTHF solvent with BS1

$\Delta G_{BS1/MeTHF+D3}$ Free energy corrected for 2-MeTHF and dispersion effects with BS1

ΔG_{MeTHF} Free energy corrected for basis set (BS2), dispersion effects and 2-MeTHF solvent

Table 29. Computed relative energies (kcal/mol) for the reaction of 3-phenyl-2-oxazolidinone (3b) at [Ru(OAc)(*p*-cymene)]. All energies are quoted relative to 3b at 0.0 kcal/mol.

	ΔE_{BS1}	ΔH_{BS1}	ΔG_{BS1}	$\Delta G_{BS1/MeTHF}$	$\Delta G_{BS1/MeTHF+D3}$	ΔE_{BS2}	ΔG_{MeTHF}
C-H Activation							
^{3b}A	0.0	0.0	0.0	0.0	0.0	0.0	**0.0**
^{3b}TS(A-B)1	21.1	20.5	22.8	21.7	14.8	21.0	**14.7**
^{3b}INT(A-B)	19.5	18.9	21.3	19.6	11.0	19.0	**10.5**
^{3b}TS(A-B)2	21.2	18.4	21.9	21.8	14.7	21.9	**15.4**
^{3b}B	14.9	14.9	18.2	14.5	10.3	15.9	**11.3**

Table 30. Computed relative energies (kcal/mol) for the reaction of 3-phenyl-2-oxazolidinone (**1ar**) at [Ru(OAc)(*p*-cymene)]. Data in bold are those used in the main text. All energies are quoted relative to **1arA** at 0.0 kcal/mol.

	ΔE_{BS1}	ΔH_{BS1}	ΔG_{BS1}	$\Delta G_{BS1/MeTHF}$	$\Delta G_{BS1/MeTHF+D3}$	ΔE_{BS2}	ΔG_{MeTHF}
C-H Activation							
Major isomer							
^{1ar}**A**	0.0	0.0	0.0	0.0	0.0	0.0	**0.0**
^{1ar}**TS(A-B)1**	20.0	19.0	18.3	16.0	14.8	18.0	**12.7**
^{1ar}**INT(A-B)**	16.6	15.4	17.2	16.3	7.4	18.9	**9.7**
^{1ar}**TS(A-B)2**	15.7	13.2	16.0	15.9	6.8	18.6	**9.7**
^{1ar}**B**	6.0	6.0	8.4	9.2	-0.3	8.6	**2.3**
Minor isomer							
^{1ar}**A′**	0.6	0.6	0.4	0.3	0.3	0.7	**0.4**
^{1ar}**TS(A-B)1′**	22.3	21.7	24.1	22.5	13.2	22.7	**13.7**
^{1ar}**INT(A-B)′**	19.0	17.8	19.5	18.0	9.1	19.7	**9.9**
^{1ar}**TS(A-B)2′**	19.2	16.9	19.9	19.0	10.2	20.4	**11.4**
^{1ar}**B′**	10.7	10.5	12.3	12.2	3.1	11.7	**4.1**

In order to estimate the Gibbs energy change for the uncatalysed coupling of 1a to 2b to give 3b, the oxidation process also needs to be included (Equation 40). The overall reaction may be considered as:

$$1a + 2b + \tfrac{1}{2}O2 \rightarrow 3b + H2O \qquad (40)$$

for which the exergonic Gibbs energy change (at 25 °C) is -49 kcal mol^{-1}. This is the amount by which all the Gibbs energies relative to A must be reduced between one turnover of the catalytic cycle and the next. (Note that the reaction path for the conversion from E2trans to A in the next cycle has not been investigated computationally.) The significance of this consideration is that it allows the turnover-dependent intermediate and the turnover-dependent transition state to be identified

as D22,1 and TS(D2-E2)1a, respectively, which occur sequentially within the same turnover cycle (Figure 36); the possibility of the turnover-dependent transition state occurring in the subsequent cycle can be discounted in this case.

Thus the computational modelling predicts the rate-determining step (as commonly understood) to be formation of the agostic intermediate immediately prior to β-hydride transfer. Under the experimental conditions the oxidation step is undoubtedly mediated by $Cu(OAc)_2$, although the detailed mechanism is unknown. However, for the present purpose it is necessary only to consider overall stoichiometry and thermochemistry, not kinetics and mechanism for this stage of the turnover cycle. The calculated free energy changes ΔG_{MeTHF} reported above (Figures 29-33) refer to a standard state of 1 atm for all species, but de-tailed considerations of the rate-determining step (or, in general, of the turnover-dependent intermediate and transition state within a steady-state catalytic cycle) depend upon actual concentrations under experimental conditions. The relative Gibbs energies shown in Figure 36 are corrected for the change from 1 atm (concentration c1) to the standard reaction conditions (Table 31; concentration c2) by the term RTln(c1/c2) at temperature T = 120 °C. The ligand exchange step B → C (for which a transition structure has not been determined) becomes exergonic due to the larger relative concentration of alkene 2b, with the consequence that the free energies of the transition structures for the alkene insertion and β-hydride transfer sections of the cycle are lowered with respect to those in the C-H activation section. The resulting profile (Figure 36) shows four transition structures (TS(A-B)1, TS(A-B)2, TS(C2-D2)2,1 and TS(D2-E2)1a) all with quite similar free energies.

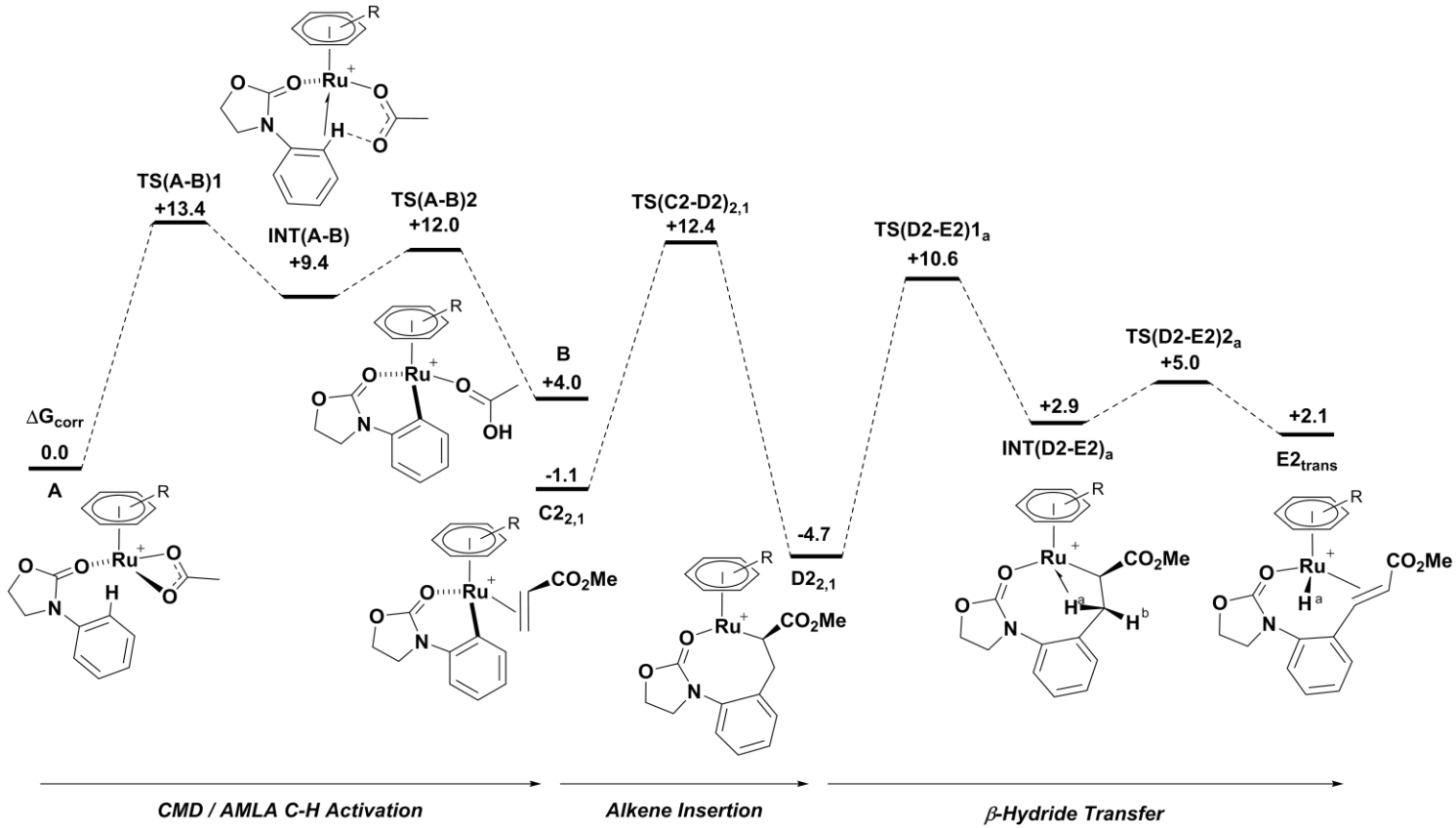
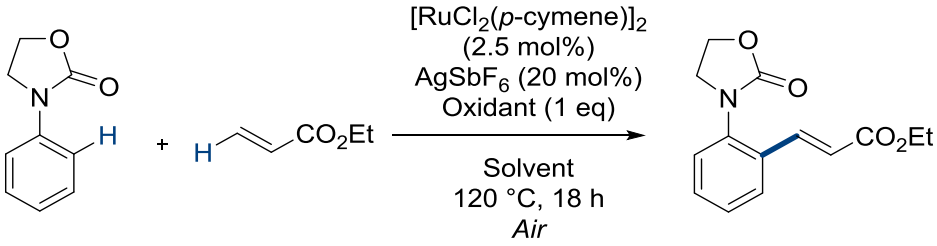


Figure 36: DFT calculated free energies (kcal mol^{-1} relative to **A**) for the coupling of **1a** and **2b** to form **3b** using a ruthenium catalyst at 120 °C in Me-THF and at concentrations corresponding to the standard experimental conditions. Graphic from J. Leitch, P. B. Wilson, et al., ACS Catalysis , 2016, 6, pp 5520-5529.

Table 31: Conversions to 3a through optimization of Oxazolidinone Directed C-H Alkenylation[a]. Where [a]General Reaction Conditions: 1a – 1 mmol, 2a – 3 mmol, [$RuCl_2$(*p*-cymene)] (2.5 mol%), $AgSbF_6$ (20 mol%), Solvent (4 mL). [b] ^1H NMR conversion. [c]AgOAc (2 equivalents). [d]Without [$RuCl_2$(*p*-cymene)]$_2$. [e]Without $AgSbF_6$. [f]Reaction performed at 100 °C [g]$AgSbF_6$ (10 mol%). [h]Solvent (1 mL). Graphic from J. Leitch, P. B. Wilson, *et al.*, *ACS Catalysis* , 2016, *6, pp 5520-5529.*



1a **2a** **3a**

Entry	Solvent	Oxidant	**3a**[b]
1	DCE	$Cu(OAc)_2$	15
2	1,4-dioxane	$Cu(OAc)_2$	56
3	DME	$Cu(OAc)_2$	72 (68)[c]
4	2-MeTHF	$Cu(OAc)_2$	70 (67)[c]
5	2-MeTHF/AcOH 3:1	$Cu(OAc)_2$	-
6	2-MeTHF	Ag_2CO_3	13
7	2-MeTHF	AgOAc	48
8	2-MeTHF	AgOAc[d]	64
9	2-MeTHF	AgO_2CCF_3	Trace
10	2-MeTHF	$Cu(OAc)_2 \cdot H_2O$	77
11[e]	2-MeTHF	$Cu(OAc)_2 \cdot H_2O$	-
12[f]	2-MeTHF	$Cu(OAc)_2 \cdot H_2O$	-
13[g]	2-MeTHF	$Cu(OAc)_2 \cdot H_2O$	68
14[h]	2-MeTHF	$Cu(OAc)_2 \cdot H_2O$	78
15[h,i]	**2-MeTHF**	**$Cu(OAc)_2 \cdot H_2O$**	**80 (76)**[c]

8.4 Kinetic Isotope Effects

Residual translational and rotational contributions to Hessian elements (punch=derivatives) were removed by a projection method and kinetic isotope effects at 393.15 K were determined within the rigid-rotor/harmonic approximation without scaling of vibrational frequencies, which satisfied the Teller-Redlich product rule; tunneling was treated by means of Bell's model applied to the imaginary frequency for motion along the reaction coordinate.

Relative to A, the computational modelling predicts an intrinsic KIE k_{H5}/k_{D5} = 2.2 for the C-H activation step. Since the preceding transition structure TS(A-B)1 for formation of the agostic intermediate is calculated to be slightly higher in energy, the intrinsic KIE would be partially masked, leading to a reduced value for the observed isotope effect. The experimental KIE by direct comparison of initial reaction rates for 1a and [D]5-1a has a value of about 2, which seems to suggest that most of the intrinsic KIE is being expressed. The KIEs calculated for the individual agostic formation, migratory insertion, and β-hydride elimination steps are all essentially unity as expected (Figure 29).

The KIE determined by the method of intermolecular competition expresses the isotopic discrimination up to the first irreversible step of the cycle, relative to free starting material.

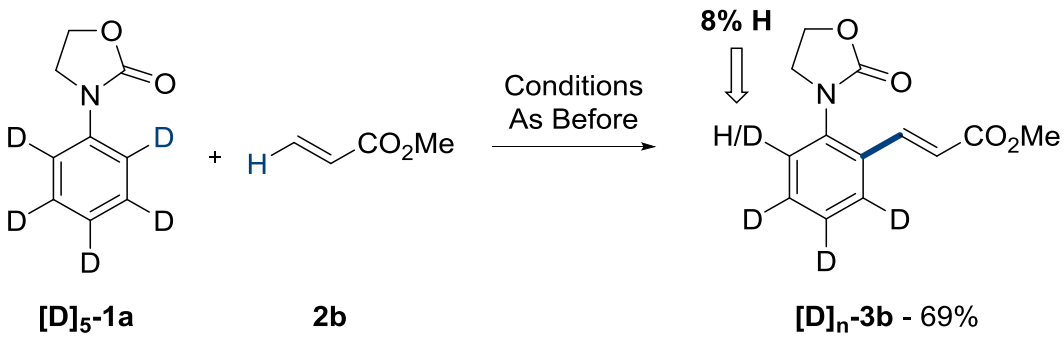


Figure 37. Ru(II)-Catalyzed Alkenylation of Isotopically Labelled Substrate. Graphic from J. Leitch, P. B. Wilson, *et al.*, *ACS Catalysis* , 2016*, 6, pp 5520-5529.*

It is known that cyclometallation is reversible (Figure 37), leading to loss of deuterium from the ortho position of 1a. The product ratio determined in the competition experiment therefore reflects the isotope effect on cyclometallation. However, the interpretation of the experimental isotope effect is complicated by at least two factors. First, the isotopic product ratio observed depends upon the degree of conversion, since starting from a 1:1 mixture of isotopologues, a product ratio of 1:1 must be obtained for 100% conversion: the observed isotope effect should be corrected for the fractional degree of conversion, which leads to a greater value for the KIE.

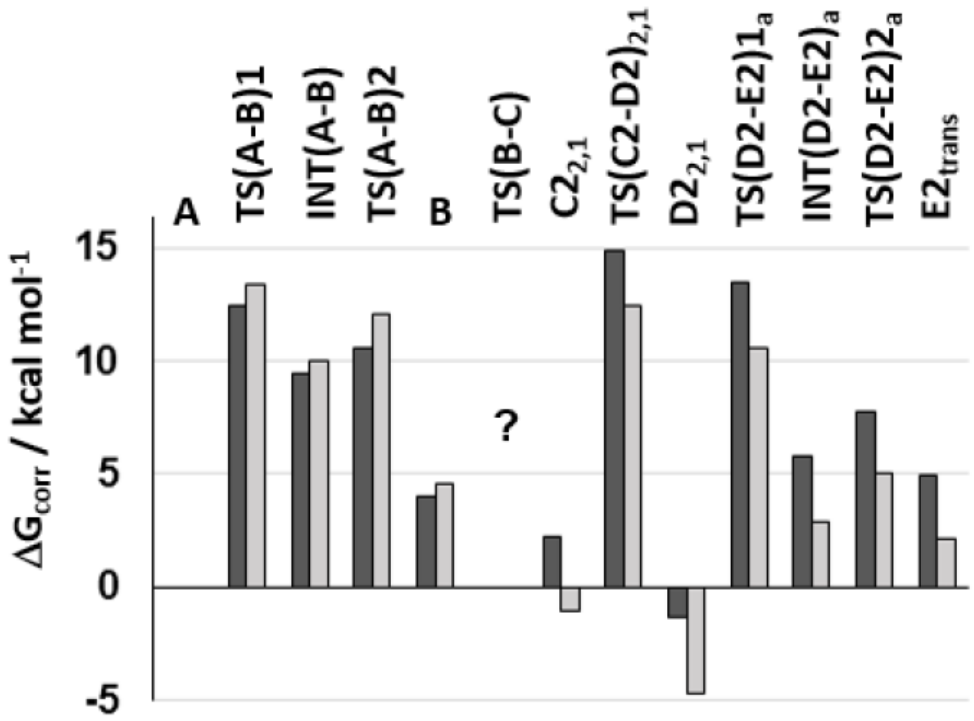


Figure 38. Relative Gibbs energies (kcal mol-1) for species in the catalytic cycle at 25 °C and equal concentrations equivalent to 1 atm (dark gray) and at 120 °C and concentrations equivalent to the actual initial experimental conditions (light gray). Graphic from J. Leitch, P. B. Wilson, *et al.*, *ACS Catalysis* , 2016*, 6, pp 5520-5529* (SI).

Second, the fact that isotopic exchange at the ortho position is seen to take place during cyclometallation means that the proportion of deuterated 1a in the reaction mixture is reduced, leading to an increase in the apparent value of the KIE. Numerical simulation shows that only a small percentage of isotopic exchange due to B → A reversibility is required to account for an apparent KIE of about 2.6 as observed for the intermolecular competition experiment.

Referring back to the novel ideas described earlier in this chapter, Figure 38 shows the Gibbs energies across an entire cycle. This incorporating the idea of turnover, would lead to subsequent cycles falling sequentially lower in energy, until completion. It is important to highlight the inherent applicability and accuracy of the protocols implemented for this complex reaction, in terms of calculation and interpretation of the isotope effects from the theoretical investigations. Of course, it is not expected that theory and experiment naturally match exactly, however the inherent similarity is encouraging, as provides further evidence that the theory and implementation of the SULISO suite is relevant to entirely different reaction scenarios than the enzymic and model systems treated in previous chapters. The predicted isotope effect matched well with experiment, and allowed for determination of the *rate determining state* and turnover parameters, as described in section 8.2.

8.5 Dispersion effects

Chapter 7 raised the interest in dispersion as an additional consideration in obtaining reliable isotope effects data from calculations. Indeed, incorporating dispersion within the optimization routine has been found to result in profound differences in geometry and therefore inter and intra-molecular interactions.[8] Figures 39 and 40 show the difference in geometry of A between dispersion-corrected, and dispersion-optimised forms.

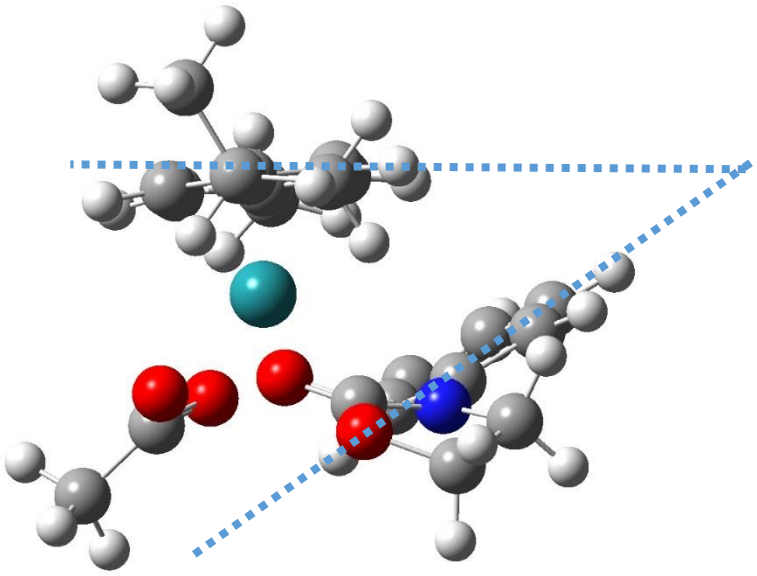


Figure 39. Dispersion-optimised geometry of intermediate A, with ring planes denoted by blue dotted lines.

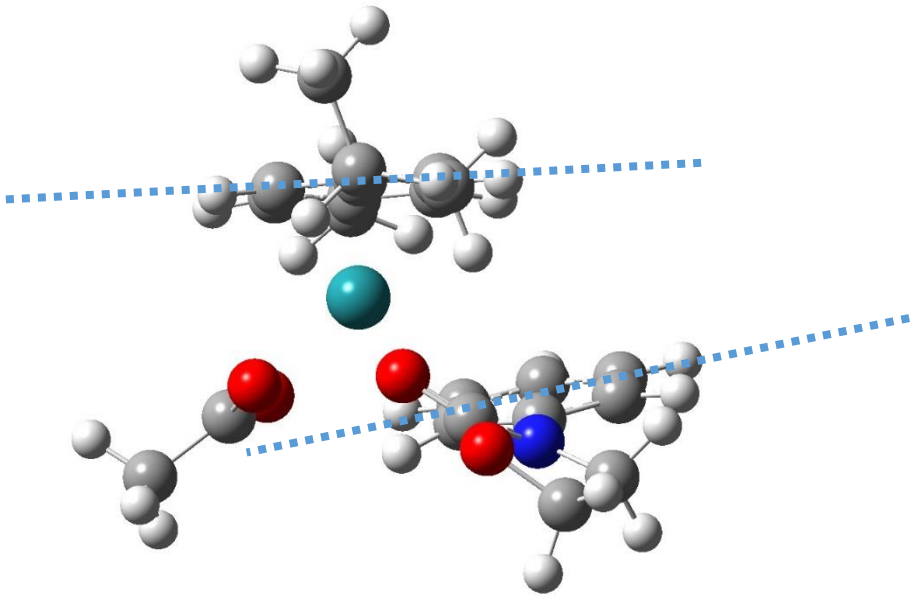


Figure 40. Dispersion-corrected geometry of A, with ring plane angles denoted by blue dotted lines.

Although there are natural, inherent similarities between the structures in Figures 39 and 40, a distinctive interaction between the ring systems in the dispersion-optimised structure are observed, compared to the dispersion-corrected equivalent. To reiterate, dispersion-optimised, or DO structures arise from incorporating the dispersion element within the optimization routine within Gaussian09. Conversely, the dispersion-corrected, or DC structures, are output from SPE calculations on the non-dispersively-corrected geometries, in order to obtain energies with incorporated dispersion.

However, the question remains of what this means to the present study. In earlier sections of Chapter 7, in was seen that incorporating dispersion has little effect on the isotope effect obtained for COMT-catalysed methyl transfer, but to what extent is this true for this ruthenium-catalysed CH-activation?

The Figures above detail the importance of dispersion within the optimization routine, leading to different geometries than DC calculations, thereby producing different Hessian matrices for calculation of the isotope effects within the reaction. In order to provide a definitive and continuous investigation of this effect, the agostic RS and TS, and CH activation RS and TS structures were recomputed with the DO protocol.

Table 32. Kinetic isotope effects for selected dispersion-corrected (DC) and dispersion-optimised (DO) reactions within the catalytic cycle in Figure 29.

	KIE (DC)	KIE (DO)
Agostic formation [A --> INT(A-B)]	1.02	2.11
C-H activation [INT(A-B) --> B]	2.17	2.87

Conversely to the work on COMT-catalysed methyl transfer, here a significant change represented by the dispersion-optimised results on the isotope effects is noted. Indeed, with the KIE for agostic TS formation raising to more than 2, this correlates well with previous work on similar reactions, where agostic formation was shown to be a major contributor in the CH activation catalytic cycle.[266, 289] However, can these changes be attributed to any particular parameter within the optimisations, or in fact a particular degree of freedom? In order to consider this, the changes in bond lengths, angles and dihedrals for the agostic formation TS were considered for the DC and DO structures.

Table 33. Selected structural parameters for structures of TS A, from which the isotope effects in line 1 of Table 32 arise. Dispersion-corrected (DC) and dispersion-optimised (DO) values are included, with a difference also quoted.

	(DC)	(DO)	Δ
average Ru-phenyloxazolidinone ring distance	2.69	2.59	0.1
average p-cymene - phenyloxazolidinone ring distance	3.57	3.44	0.13
p-cymene - phenyloxazolidinone ring plane angles	-55.1	-48.4	6.7

In order to consider the basis by which the DO KIEs quoted in Table 32 vary from the DC results, the values in Table 33 were tabulated. Key parameters which have previously been described as affecting the magnitude of isotope effects in previous chapters, include environmental influences on the sites of isotopic substitution. These being the hydrogen atoms in the phenyloxazolidinone ring, their proximity to other moieties will lead to varying electrostatic and steric effects.

One effect of dispersion incorporation as detailed by Schreiner and coworkers, includes hydrogen bonding, and electrostatic contributions.[8] Of course, in earlier chapters the marked effect electrostatics have on isotope effects and their contributing factors have been demonstrated, however dispersion incorporation brings to light the inherent sensitivity of these effects to geometrical change. Indeed, with the tilting effect of the ring in the DO structure, both the *p*-cymene pi system and phenyl system on the oxazolidinone are able to align, effectively placing the isotopically sensitive atoms on the oxazolidinone ring in closer proximity to the electronegative effects of the oxygen atoms on the acetyl fraction.

It can be expected that this more reasonably represents the structures responsible for the experimental isotope effects, where the conjugate systems would exhibit an energetic favourability to increased interaction, thereby contracting the scope of the entire structure.

8.6 Summary

Having explored an additional feature contributing to the output of calculated structures in common computational methodologies, two examples of incorporation of dispersion effects have been seen.

In the case of COMT-catalysed methyl transfer, DO structures varied little in KIE from their NDC equivalents. Indeed, the isotope effect itself did not significantly change with dispersion incorporated. Conversely, in the case of the ruthenium-catalysed C-H activation, described in the latter half of this chapter, significantly different isotope effects obtained from the DO as compared to the DC structures have been shown. Indeed, in the DO calculations, the KIE for agostic TS formation increased to 2.11, from 1.02 in the DC. This has been rationalised through the changes in geometry observed with DO.

Of course DC calculations are less time consuming, and in some cases would be sufficient (as in COMT), however for the catalytic cycle discussed in this chapter, the importance of including dispersion effects within the optimisation procedure and therefore producing a *structure* as opposed to simply an energy which has been dispersively corrected, results in markedly different KIEs.

In order to explain this, a subset of the internal coordinates were shown in Table 33, whereby increased interaction between the pi systems on the conjugated rings, and environmental electrostatic effects on the isotopically sensitive atoms were responsible for the differences observed in KIEs.

Having now described numerous effects on computational KIE and EIE calculations, it is timely to review the findings of this thesis, in terms of recommended protocols for isotope effect calculation and interpretation, which will be the subject of Chapter 9, which will go on to briefly discuss current and future work in the field, with particular emphasis on additional guidance which can be applied to the discipline at large.

9. Recommendations for isotope effect calculations in the supramolecular age, and Future Work

This thesis had the aim to provide meaningful, thorough recommendations for the future calculation and interpretation of isotope effects. The theory of KIEs and EIEs, their calculation and related parameters formed the initial chapters, moving on to examples from the literature.

This was followed by a detailed description of the computational theory behind the calculations within the thesis, as well as those described from other sources. Electronic structure theory was introduced, moving onto larger systems with QM/MM calculations and molecular dynamics.

Chapter 2 deepened the theoretical background by describing the suite of isotope effect calculation software and vibrational characterisation codes used in the context of this work. This was developed into alternative methods for isotope effect calculation and estimation of drug-related discovery in computational science.

Chapter 3 provided the first introduction into the question which would be developed throughout the thesis in its entirety; namely factors influencing isotope effects which perhaps are less traditionally considered and result in significant impacts not only on results but also on their interpretation. The methyl cation was shown to exhibit significant isotopic sensitivity when it came to solvent environment and electrostatic interactions, suggesting that similar considerations should be taken into account for general isotope effect calculation.

Chapter 4 provided a more rounded methodological study on the impact of different electronic structure methods on the vibrational frequencies and isotope effects obtained from the methyl cation and its isotopologues. In addition, it was recommended that caution be used when combining anharmonic corrections with certain DFT functionals (including M06), as well as the PCM parameterisation in Gaussian09, which produces anomalous results. The major finding of this work, was that anharmonic corrections being computationally expensive, are not necessary form small basis set calculations in order to obtain more reliable vibrational frequencies. Instead, scaled harmonic frequencies can provide the accuracy equivalent to anharmonic corrections, and in some cases, superior to these.

With Chapter 5, the discussion moved from implicit solvation to explicit, considering the inherent effect solvent distance and shell formation will have on the isotopic sensitivity of a system. This included an additional investigation into the nature of the degrees of freedom within a system, with our novel approach of a superheavy cage, showing the effect spurious motions of solvent atoms have on the reliability of isotope effects obtained from calculations. By tethering these motions as in a protein environment, isotope effects for the full system reflected a not insignificant effect from environmental effects perpendicular to the reaction coordinate. Indeed, it was shown that these interactions must be taken into account for a system in order to provide a thorough and balanced interpretation of isotopic sensitivity and the system interactions.

The cutoff procedure was introduced in more detail within Chapters 6 and 7, with the cage system being the subject of Chapter 5, being used as the test system for application of cutoff. It was noted both for the cage system and QM/MM structures that initial 3-bond cutoffs as suggested by Wolfsberg and Stern did tend to produce reasonable results, although cutoffs affecting the reaction coordinate significantly resulted in marked deterioration of the reliability of the isotope effect obtained.

Chapter 7 considered another factor in common electronic structure calculations, this being dispersion. Indeed, larger systems often include dispersion interactions as these have historically been seen to be significant in magnitude only for these, however recent work has shown that dispersion can have a significant effect for smaller systems also, in terms of electrostatic and conjugative properties. It was shown that for the COMT-catalysed methyl transfer investigated in the context of this chapter, dispersion effects did not result in significant structural or isotopic change.

Chapter 8 moved on to describe work carried out in collaboration with synthetic organic colleagues, on ruthenium-catalysed C-H activation. This was important for the methodology used within our isotope effect studies, as proved the validity and inherent reliability of our theory in comparison to the experimentally-obtained results. Additionally, this allowed us to again consider the effects of dispersion, which were markedly more significant here than in the COMT work.

Indeed, including dispersion within the optimisation routine as compared to in a correction within a single point energy calculation, had the result of altering the geometry and subsequently the isotope effects obtained for steps in the cycle. This was explained through the change in directionality of the ring plane interactions, with additional pi-stacking from each conjugated ring occurring in the DO structure as opposed to the DC structure.

It is clear that a number of points are worthy of note from this work, each to be explicitly considered not only in the calculation, but also interpretation of isotope effects. The following guidelines are structured in order to serve as recommendations for researchers using any combination of electronic structure code. Examples are given of codes (such as SULISO for isotope effect evaluation) however other examples exist and are mentioned appropriately. The recommendations are broadly centred around the conclusions from each chapter of this thesis, in order to provide some continuity to the presented work.

9. 1. Solvent effects on isotope effects not only exist, but can have a profound influence on KIEs and EIEs

The effect of the environment on computed isotope effects has been the subject of some discussion throughout this work, and the concept of solvent effects on isotope effects features heavily. It has been shown that the equilibrium isotope effect for methyl transfer from vacuum to a PCM solvent varies significantly with the dielectric constant of the solvent.[1] In particular, this variation is most pronounced within the range of dielectric constants of 1 to 10; the range of dielectrics quoted for most enzyme active sites.[290] It is therefore recommended that care be taken when interpreting mechanistic data from enzyme catalysed reactions in solution, in order to ensure that calculated or observed isotope effect, or their variations between mutants, is not simply an artefact of the changes in dielectric environment that the substrate is exposed to. It is also recommended that calculations take place for species in enzyme and solution, in order to gauge solvent effects (if any) on isotope effects for the reactions.

9.2. Anharmonic corrections are not necessary for electronic structure methods including small basis sets

Anharmonic corrections can be computationally expensive to implement into electronic structure calculations, therefore justification for their inclusion must be made.[189] It has been pointed out,[4] that certainly for electronic structure methods (ESMs) with small basis sets, it is not necessary to include anharmonic corrections within the optimisation or energy evaluation procedure, instead opting for empirical scaling factors, which effectively capture the influence of anharmonicity and additional effects. Anharmonic corrections as implemented within Gaussian09 under the Barone scheme[194, 195] seem to be highly method dependant in their reliability; combinations of these corrections with the M06 functional produce erroneous results for vibrational spectra of the methyl cation isotopologues. In addition, unusual results are obtained from combinations of B3LYP with three basis sets of increasing size, with anharmonic corrections under the polarised continuum model (PCM). The reasonable data obtained from previous B3LYP calculations under the harmonic approximation within the PCM[1] suggest that it is the combination of these implementations which leads to erroneous data for the methyl cation and its isotopologues. The fact that the substrate is cationic in nature has also been suggested to be a possible cause of the breakdown of these implementations when combined.[291]

Of course, post-HF methods such as CCSD include anharmonic effects implicitly within their treatment of electronic structure, however with the prevalence of DFT methods as computationally cheap and often reasonable alternatives to the more expensive correlation methods, the question of the benefits of implementing anharmonic corrections remains. Indeed, certainly for some systems, anharmonic corrections have been shown to greatly improve the accuracy of results relative to experiment[197], however in the context of these investigations, the robustness of current methods remains questionable. As such, it is recommended that for calculations with small basis sets (6-31+G(d), or similar), scaling factors be employed instead for the treatment of vibrational frequencies. For larger basis sets and problems requiring the implementation of anharmonic corrections, certainly as applied within Gaussian09, it is recommended to perform both harmonic and anharmonic calculations, and carefully compare the results obtained to experimental, or higher-level computational results before forming definitive conclusions. As we have seen, certain DFT functionals such as M06 do not perform well in combination with these anharmonic corrections within Gaussian09,

producing erroneous vibrational frequency, therefore care must be taken when interpreting and analysing results. Additionally, the use of anharmonic corrections within Gaussian09 is not recommended in combination with the PCM, certainly for charged species, unless initial benchmarking has taken place. It has been shown that the combination of these implementations leads to erroneous vibrational data for the B3LYP functional and 6-31+G(d) basis set, which alternatively perform well under the harmonic approximation when compared to experiment. Although care must be taken to implement anharmonic corrections within calculations in Gaussian09, there is no doubt that inclusion of anharmonic effects is likely to become commonplace with increases in computational performance allowing for anharmonic corrections to be readily implemented within DFT, or more frequent use of CCSD and other electron correlation methods. It is also important to highlight the recent release of Gaussian16, the latest implementation of the popular electronic structure code. It has been suggested that many routines have been overhauled, therefore the issues causing the erroneous vibrational results obtained for the methyl cation isotopologues with Gaussian09 may be solved within Gaussian16. An investigation of this is currently underway.

9.3. Environmental effects can influence isotope effects in solution and enzyme

The discussion on solvent effects broadly led to a more in depth investigation into explicitly solvated systems, and the effect of influences other than those in the direction of a reaction coordinate, on vibrational data and isotope effects for a reaction. The methyl transfer within a constrained cage of five water molecules at the vertices of a trigonal bipyramid was considered as a model system for COMT catalysed methyl transfer. With the linear positions of the vertices of the trigonal bipyramid being readily variable, it was possible to effectively vary the level of solvation of the methyl cation in the centre, by varying the distance of the so-called *axial* and *equatorial* water molecules within the cage structure. Condensed structures with small distances between the methyl cation and the surrounding water molecules replicated highly solvated systems with exhibiting large values of dielectric constant such as enzyme active site. Conversely, expanded cage structures with large distances between the methyl cation and each vertex of the trigonal bipyramid modelled weakly solvated systems exhibiting low dielectric constants.

Changes in both the methyl transfer axis, or *axial* direction, and that perpendicular to the reaction coordinate, or *equatorial* direction led to variations in the calculated isotope effects. Indeed, variation in the axial direction produced significant changes to the observed isotope effect, leading to a change from S_N2 methyl transfer at short axial C_{methyl}-$O_{axial\text{-}water}$ distances, to a more S_N1-like process at larger axial extensions. This mechanistic change is noteworthy as highlights the sensitivity of mechanism to the geometry of the reaction coordinate. Perhaps less predictable however of some significance, is the effect of the equatorial substituents, and their distances from the methyl cation, on the calculated isotope effects. These substituents form hydrogen bonds with the hydrogen atoms on the methyl cation, and are directly positioned at each vertex of the C-H bonds. Although smaller in magnitude to the isotope effect differences observed in changing axial substituent distances, the effects of changing the distances for these hydrogen-bonded moieties were nonetheless significant. Moreover, earlier work by Klinman and coworkers had considered changes in KIEs within the COMT active site, between WT and mutant COMT, to support the compression hypothesis, a suggestion that the active site mechanically forces reactions to occur, leading to its catalytic power. However, the data presented by Klinman, where 2^o-α T_3 KIEs on k_{cat}/K_M were found to be 0.791 ± 0.012 for wild type human COMT, and 0.822 ± 0.021 and 0.850 ± 0.012 for the Y68F and Y68A mutants are readily explained instead by the changes in environmental effects perpendicular to the reaction coordinate, rather than axial compression. Indeed, $k(CH_3)/k(CT_3)$ for methyl transfer within the constrained cage of water molecules becomes *less inverse* by 3% when interactions perpendicular to the reaction coordinate increase in magnitude (0.5 decrease in hydrogen bonding distance). This suggests that at least partially, observed KIEs for mutant structures of COMT as compared to the WT can be rationalised by changes in the local dielectric of the enzyme active site. This provides meaningful evidence of an electrostatic origin to catalysis in AdoMet-dependant methyltransferases. This also suggests that for future interpretations of isotope effects in enzymes, the effect of the local environment or changes therein must be considered in order to ensure that accurate and reliable conclusions are drawn from the isotope effects and associated data. This can be accomplished by discerning the effects of local residues present in the active site on the isotope effect individually, and in combination, as well as constructing model systems to identify the origin of the electrostatic contribution to catalysis.

9.4. The vibrational origin of isotope effects is important and needs to be identified for reliable interpretation

Isotope effects depend on a number of components. From the statistical mechanical and transition state theory approaches, it is understood that three main components contribute to the magnitude and direction of isotope effects.[2] The vibrational excitation term, or EXC as referred to throughout this thesis accounts for excited vibrational states and the effect population of these will have on an isotope effect. The mass and moments of inertia term takes the molecule as a rigid rotor and includes contributions from the translational and rotational energies. Finally, the zero point term, or ZPE, describes the contribution to the isotope effect from the energies of 3N-6 normal vibrations. In each reaction or problem in which isotope effects are employed to probe the nature of the reactivity and mechanism, different vibrational contributions will dominate in the determination of the isotope effect. It is often the case for ground state systems that the ZPE component will dominate. The question however must be asked as to which vibrational modes contribute significantly to this component to the IPFR and thus isotope effect. Isotopically-sensitive stretches and bends of the isotopically substituted atoms are often identified as principally contributing to the magnitude or trends in ZPE, however as noted earlier,[3] other modes can contribute as effectively. Indeed, although the trend in isotope effect with distance of equatorial hydrogen-bonding moieties was dominated by the C-H stretching vibrations in the methyl cation in Reference 3, the overall normal direction of the isotope effects was governed by bending modes and external degrees of freedom (translations and rotations). Inspection of these components is often of great value in successfully interpreting isotope effect data, as indicates the isotopic sensitivity, and contributions of the atoms within a chemical system. It is recommended that this approach be taken when interpreting isotope effects and vibrational data, in order to ensure that conclusions match the underlying vibrational data obtained. Additionally, inspection of these components can be vital in assessing the reasonableness of the data obtained. Indeed, due to the nature of embedded systems such as the methyl cation with the constrained cage of water molecules, which simply models a larger QM/MM system, vibrations of the reacting system can become inherently coupled with the environmental atoms. This was observed previously, and led to the use of the so-called *superheavy approach*. In coupled systems, it is often the smaller atoms and moieties which can lead to slight errors in vibrational calculations due to spurious vibrations. In order to avoid these

whilst still maintaining the use of the same system, the masses of these smaller moieties can be modified to large magnitudes in order to effectively tether these in space and avoid unusual vibrational results. This can easily be accomplished with the SULISO software,[292] where the masses of atoms can be edited in the standard input. Alternatively, other isotope effect calculation software allows for this, also.[17] Another advantage to this technique, is the mimicry of a supramolecular system. Indeed, COMT has a mass of approximately 30 kDa, and with each hydrogen on water molecules in the constrained cage described in Reference 3, replaced by atoms of mass 999, the mass of the constrained cage becomes >10kDa, a figure far closer to that of the environment of the modelled reaction, than previously. Of course, it is not suggested that this provides a viable alternative to large scale computations such as QM/MM simulations, as this mass effect certainly does not take into account the electrostatic, steric and additional effects afforded by state of the art QM/MM techniques.

9.5. The use of the cutoff approximation is both valid and reasonable in computational studies of large systems

The subject of some scepticism, the cutoff approximation, where a molecular Hessian matrix is truncated to rows and columns corresponding only to a subset of the atoms in the system, remains both usable and reliable when applied judiciously.[224] Cutoff methods are applied as often the full optimised system is not used in computation of KIEs from large QM/MM structures, instead a subset of the supramolecular system such as the QM region, or smaller, is used to compute KIEs. As such, the implementation and manner in which the system is truncated are particularly important in correctly and reliably evaluating isotope effects arising from analysis of large structures. Indeed, it is recommended that, particularly when truncated Hessian matrices for KIE computation, the description of the reaction coordinate remains the same as in the original system. It has been found that cutoffs particularly involving truncation of atoms or residues along the reaction coordinate lead to large errors in computed isotope effects compared to the full system descriptions. This by should be no means dissuade users of program CUTOFF (as included in the SULISO suite), or other software, however it is important to ensure that reasonable cutoffs are established based on the reacting system. For example, Wolfsberg and Stern originally suggested two, or three bond cutoffs (where the number of bonds is taken from the isotopically substituted atom as the origin).[15, 227]

It has been noticed herein that three-bond cutoffs can provide reasonable models of the full systems, however ensuring the accurate description of the reaction coordinate, especially for KIEs remains important. This is accomplished by inspecting the modes and atoms which are isotopically sensitive and contribute to the total KIE, ensuring that these remain in the truncated system. It has also been noted that computation of EIEs is less sensitive to breakdown in the description of the reaction coordinate than KIEs. Indeed, the components of EIEs tend to be conserved with more truncated systems, largely due to their dependence more on environmental effects than an explicit coordinate as in KIEs. It is also recommended to treat isotope effects with the so-called *all frequencies* method, where small imaginary vibrations are treated as real. This has been found to lead to more accurate representations of systems, rather than treating these as alternative motions such as librations. The SULISO suite allows for designation of imaginary modes, or their treatment as real vibrations through the modification of the LIPFR and UJISO codes. Users are directed towards the manuals for these pieces of software, which are available on the suite GitHub archive: https://github.com/pbw20/SULISO

9.6. Dispersion corrections should be called within the optimisation procedure and not only for single-point energy calculations

Dispersion interactions, being the attractive portion of the van der Waals potential have been the subject of some discussion based on their implementation in common electronic structure methods. Some higher level techniques such as MP2 and CCSD include these implicitly with their perturbative or correlated treatment of electrons, however most DFT methods do not. As such, the question of where and more importantly when to apply dispersion corrections remains relatively unanswered. The implementation of dispersion corrections for the cage test system, COMT catalysed methyl transfer to catecholate, and Ruthenium catalysed C-H activation reactions were recently tested for computational expense based on employing dispersion within the optimisation, thereby leading to different geometries than implementing the correction simply as a post-optimisation single point correction. Very little difference in the timescale of simulations was found, leading to the suggestion that the incorporation of dispersion corrections within the optimisation routine is not onerous. This is recommended for systems where dispersion is likely to have an effect, and due to the relative computational ease of implementation, could be included in all calculations. It has been shown that dispersion corrections

implemented within the optimisation (DO) often lead to structures which differ in geometry to the non DO equivalents. This naturally has an effect on the energetics but can also have a profound effect on other properties, such as the isotope effects computed for reactions. Indeed, it has been demonstrated that for some reactions, DO has little effect on the KIEs, and some effect on the computed reaction barriers. Conversely, for the ruthenium-catalysed C-H activation study, DO structures lead to largely different computed KIEs for the reactions. It is therefore recommended that dispersion corrections are implemented within optimisation calculations, and not just as post-optimisation single point corrections. This of course can lead to the same results, however it has been shown that at least for some reactions, DO geometries lead to different KIEs. In the context of this work, dispersion has been implemented within Gaussian09, using Grimme's D3 correction with Becke-Johnson damping, however a number of alternatives also exist, and the reader is directed to a recent review by Wagner and Schreiner, which details the main points and common codes.[8]

9.7. Toolkit for supramolecular isotope effect calculations

A brief summary of the recommendations for supramolecular KIE and EIE calculations is included below, and intended for use as a broad-spectrum set of guidelines:

- Correctly describing the reacting system is vital in terms of isotope effects: all (or the most significantly) isotopically-sensitive vibrational modes must be taken into account when computing isotope effects, as omitting these will lead to errors.
- The description of the system is carried out through an initial optimisation (QM or QM/MM most likely). This should equally reflect the real system as far as possible. Electrostatics can be very important in enzyme catalysis and a correct description of these, particularly in the active site, will lead to more reliable results.
- Dispersion corrections should also be employed during the optimisation procedure. These are not computationally expensive, and will likely improve the reliability of the geometries obtained from optimisation routines. This will in turn provide more reliable isotope effects.
- Anharmonic effects can affect results, however are included in MP2 and higher-level computations. In QM/MM systems where the size of system is

prohibitively large for such calculations, the use of scaling factors for DFT methods is suggested. Currently, there are difficulties in the implementation of anharmonic corrections within Gaussian09, therefore users of this software are urged caution when combining these corrections with DFT functionals, in particular for charged systems. It is recommended that these corrections are employed and compared to the same calculation under the harmonic approximation in order to assess the reasonableness of results.

- Once reasonable structures have been optimised, it is important to correctly calculate the KIEs and/or EIEs for the reacting system. The SULISO suite[292] and the ISOEFF[17] program are two pieces of software which can accomplish this. It is recommended that small imaginary frequencies arising from spurious small atom motions within a QM/MM system are treated as real, and not ignored. This generates more reliable isotope effects.
- Reliable interpretation of isotope effects is important. When considering the origin of the isotope effects obtained, factors other than those described in the reaction coordinate must be considered. The contributions from the local electrostatic environment, sterically bulky moieties, and other environmental effects must be taken into account in order to accurately describe the nature of the isotope effects obtained. In depth analysis of the components of KIEs and EIEs is provided in the output files from the programs in the SULISO suite, where the individual contributions from atoms and individual modes to the isotope effect can be obtained. This will lead to reliable mechanistic interpretations of the nature of isotope effects.

Isotope effects, as such, are widely used as the most subtle of experimental probes, and a detailed theoretical and methodological understanding of isotopic sensitivity is inherently necessary for reliable interpretation of experimental results and calculations. Herein, recommendations and examples have been provided to supplement the field with a detailed analysis of the basis behind interactions responsible for KIEs and EIEs. In the final part of this chapter, very recent advances in the field, as well as avenues for future work will be considered.

9.8. Recently published results and future work

One aspect of isotope effect science, is that it is constantly changing due to the nature of the subject. Of course, this can be said for numerous disciplines within the scientific community, however the breadth of incorporation of isotope effect studies within general research is exceptional.

Very recently, two rival groups have published on the subject of QM region size within QM/MM calculations, and the impact of results obtained, which include isotope effects.[293, 294] This is inherently relevant to the work presented herein, as touches upon a similar discussion as that of the cutoff approximation. Indeed, in some way, the size of the QM region is the first step in a cutoff implementation and establishes the maximum size of the region for explicit calculation of isotope effects. The description of this region can thereafter be truncated by the cutoff procedure, however if the QM region itself is not large enough then the isotope effects will not be accurate reflections of the reacting system.

The first, by Jindal and Warshel highlights that obtaining results which can be considered reliable, requires significant sampling, as well as a reliable QM approach itself. They modelled the free energy profile of the reaction of COMT-catalysed methyl transfer, with convergence modelled based on the size of the QM region. As can be expected, analogous to the cutoff work in earlier chapters of this thesis, Warshel and Jindal note that the smallest subsets/ smallest QM regions, tend to yield results less representative of the entire system. Warshel notes that there are more marked differences between size of system, when calculated exothermicity is considered. They point out that systems without ionic incorporation lead to less reliable results, in keeping with work presented earlier in this thesis, relating to electrostatic and charge-related influences on systems.

Warshel and Jindal also show the importance of not only including direct reaction coordinate molecules within the QM region, with inclusion of perpendicular residues resulting in substantially greater accuracy in the activation free energies. These energies corresponding to inclusion of different residues within the system are confirmed as not being additive, an analogous situation to the isotope effects obtained within the anharmonic evaluations carried out earlier in this thesis.

In terms of COMT mutants, which are the subject of the compression controversy, Warshel and Jindal point out that a large QM region must be used, as the effects of

the mutations must be considered as a whole, rather than a subset; certainly in the first instance.

There is additional information provided by Warshel and Jindal, on the importance of correct electrostatic and structural data. Indeed, they describe the importance of a correct dielectric description of the active site cavity in order to obtain reliable results, a conclusion supported by the cage and QM/MM work included in this thesis.

Warshel and Jindal suggest that the size dependence on reliability of QM/MM system calculations in not as important as perhaps expected. Indeed, they consider that the dependence on size is far smaller than the catalytic effect of the COMT enzyme used as the test case within the study.

They highlight the importance of the donor-acceptor axis, indicating that residues within this have been shown to significantly contribute to the catalytic effect and nature of the reaction. Additionally, they comment on the environmental residues, suggesting that although exhibiting less of an effect on the substrate and reaction, their inclusion contributes to the dielectric environment and steric effects of the system. They conclude by stating that size dependence may not be as critical as previously thought; including residues not in direct contact with the reaction coordinate (*i.e.* similar to our three bond cutoff) does not change the activation barrier in a significant way.

Klinman and coworkers also recently treated the size dependence of QM/MM calculations using the COMT enzyme as an example. They considered systems from 64 atoms, through to 940 atoms within the QM region. They show that structural and chemical properties of COMT slowly approach the asymptotic limits, as they added residues to the QM region.

Klinman and coworkers described how distances between reactants and partial charges on these, converged slowly with increasing QM region size, highlighting that through the use of multiple functionals, they had limited the effect of correlation-related error responsible for some previously quoted results.

They described how their large QM/MM simulations showed that the charge transfer process within the reaction, does not occur until after the TS, and that this charge transfer strongly affects the orientation of the ES complex. The group introduce a

concept known as charge-shift analysis, in identifying the residues key to obtaining reliable results. It is interesting to note that this produced very good results for example in using only 16 residues out of 56 in a system, however some of the 16 would not have been chosen from proximity or chemical intuition, suggesting that their charge/electrostatics based choice of residues was important.

Klinman and coworkers conclude that forces on the central QM atoms do not converge until the QM region reaches more than 500 atoms, a suggestion which remains to be tested for KIEs, which would be affected by the correct description of these forces. Indeed, it is suggested that QM regions of more than 10 times the typical size be used in the future for reliable structural or energetic evaluations, however for simple quantitative accuracy, it is possible to carry out the charge shift analyses as described by Klinman and colleagues, and obtain reasonable data from only 16 residues in the active site.

There is some disagreement between these published results. Indeed, Warshel and Jindal suggest that, other than including residues directly in contact with the reaction coordinate, the accuracy of computed activation barriers does not change significantly when including additional residues. Conversely, Klinman and coworkers mainly claim that for reliable structural and energetic evaluations of structures from QM/MM calculations, QM regions must be approximately 10 times their current typical size, and for converged forces in the active site, more than 500 atoms must be included in the QM region.

9.81. Future Work

Future work in this field will build upon the guidelines established in this thesis to develop our understanding of additional factors influencing the reliable calculation of KIEs and EIEs in supramolecular systems. Work is currently being undertaken by the author to verify the published anharmonic data by computations in the thoroughly revised Gaussian16 program.

Additionally, three publications are in preparation. One treats the S_N2 to S_N1-like change in mechanism observed for the cage structures described in Chapter 5, and the factors influencing this. Additionally, the dispersion optimised results for ruthenium-catalysed C-H activation are being prepared for publication as an article aimed at recommending the use of dispersion optimisation as opposed to the use of

dispersion as a single point correction. Finally, the cutoff results presented for the cage system are currently under preparation for a manuscript.

Work will focus on using the Warshel[295] and Klinman[293] studies to describe the effect of QM region size in QM/MM calculations, on computed isotope effects, and their reliability. This will feature as a complementary study to applications of the cutoff procedure, and will include use of the cutoff technique within the QM region sizes also, in order to reassess truncation.

As a final notion to complete this thesis, a number of recommendations have been put forward through judicious use of theory combined with experiment and higher level methodologies, in order to provide the field of isotope effect science with additional guidance in calculation and interpretation. It is the hope that this thesis will help guide future investigations, and suggest new ways to interpret isotope effects on the basis of molecular structure and chemical data obtained.

References

1. P. B. Wilson, P. J. Weaver, I. R. Greig and I. H. Williams, *J Phys Chem B*, 2015, **119**, 802-809.

2. I. H. Williams and P. B. Wilson, in *Simulating Enzyme Reactivity: Computational Methods in Enzyme Catalysis*, The Royal Society of Chemistry, 2017, DOI: 10.1039/9781782626831-00150, pp. 150-184.

3. P. B. Wilson and I. H. Williams, *Angew Chem Int Edit*, 2016, **55**, 3192-3195.

4. P. B. Wilson and I. H. Williams, *Mol Phys*, 2015, **113**, 1704-1711.

5. P. Wilson and I. Williams, *Manuscript in preparation*, 2017.

6. J. A. Leitch, P. B. Wilson, C. L. McMullin, M. F. Mahon, Y. Bhonoah, I. H. Williams and C. G. Frost, *ACS Catal*, 2016, **6**, 5520-5529.

7. J. N. Israelachvili, *Proceedings of the Royal Society of London*, 1972, **331**(1584), 39-55.

8. J. P. Wagner and P. R. Schreiner, *Angew Chem Int Edit*, 2015, **54**, 12274-12296.

9. J. Bigeleisen and M. G. Mayer, *J Chem Phys*, 1947, **15**, 261-267.

10. J. Bigeleisen and M. Wolfsberg, *Adv Chem Phys*, 1958, **1**, 15-76.

11. A. Streitwieser, *Progress in Physical Organic Chemistry*, Wiley, 2009.

12. L. C. S. Melander and W. H. Saunders, *Reaction rates of isotopic molecules*, Wiley, New York 1980.

13. J. Bigeleisen, *J Chem Phys*, 1958, **28**, 694-699.

14. M. G. Redlich, *P Natl Acad Sci USA*, 1953, **39**, 560-563.

15. M. J. Stern, W. A. Vanhook and M. Wolfsberg, *J Chem Phys*, 1963, **39**, 3179.

16. M. Saunders, K. E. Laidig and M. Wolfsberg, *J Am Chem Soc*, 1989, **111**, 8989-8994.

17. V. Anisimov and P. Paneth, *J Mathemat Chem*, 1999, **26**, 75-86.

18. S. Wolfe, S. Hoz, C. K. Kim and K. Y. Yang, *J Am Chem Soc*, 1990, **112**, 4186-4191.

19. N. Balazs, in *Part I: Physical Chemistry. Part II: Solid State Physics*, ed. A. S. Wightman, Springer Berlin Heidelberg, Berlin, Heidelberg, 1997, pp. 3-20.

20. R. P. Bell, *Trans Faraday Soc*, 1959, **55**, 1-4.

21. P. F. Cook, *Enzyme Mechanism from Isotope Effects*, Taylor & Francis, 1991.

22. J. H. Schachtschneider and R. G. Snyder, *Spectrochim Acta*, 1963, **19**, 117-168.

23. E. B. Wilson, J. C. Decius and P. C. Cross, *Molecular vibrations : the theory of infrared and Raman vibrational spectra*, Dover Publications, New York, 1980.

24. L. C. S. Melander and W. H. Saunders, *Reaction rates of isotopic molecules*, Wiley, New York, 1980.

25. W. D. Gwinn, *Journal Chem Phys*, 1971, **55**, 477-481.

26. J. Mckenna, L. B. Sims and I. H. Williams, *J Am Chem Soc*, 1981, **103**, 268-272.

27. H. L. Sellers, L. B. Sims, L. Schafer and D. E. Lewis, *J Mol Struct*, 1977, **41**, 149-151.

28. I. H. Williams and G. M. Maggiora, *J Mol Struct Theochem*, 1982, **6**, 365-378.

29. I. H. Williams, *Chem Phys Lett*, 1982, **88**, 462-466.

30. H. S. Johnston, *Gas phase reaction rate theory*, Ronald Press Co., New York, 1966.

31. G. W. Burton, L. B. Sims, J. C. Wilson and A. Fry, *J Am Chem Soc*, 1977, **99**, 3371-3379.

32. P. J. Berti, *Method Enzymol*, 1999, **308**, 355-397.

33. J. Rodgers, D. A. Femec and R. L. Schowen, *J Am Chem Soc*, 1982, **104**, 3263-3268.

34. J. D. Hermes and W. W. Cleland, *J Am Chem Soc*, 1984, **106**, 7263-7264.

35. F. Mentch, D. W. Parkin and V. L. Schramm, *Biochemistry*, 1987, **26**, 921-930.

36. D. W. Parkin, F. Mentch, G. A. Banks, B. A. Horenstein and V. L. Schramm, *Biochemistry*, 1991, **30**, 4586-4594.

37. B. A. Horenstein, D. W. Parkin, B. Estupinan and V. L. Schramm, *Biochemistry*, 1991, **30**, 10788-10795.

38. D. J. Merkler, P. C. Kline, P. Weiss and V. L. Schramm, *Biochemistry*, 1993, **32**, 12993-13001.

39. J. Scheuring and V. L. Schramm, *Biochemistry*, 1997, **36**, 4526-4534.

40. K. A. Rising and V. L. Schramm, *J Am Chem Soc*, 1997, **119**, 27-37.

41. P. J. Berti, S. R. Blanke and V. L. Schramm, *J Am Chem Soc*, 1997, **119**, 12079-12088.

42. J. Scheuring and V. L. Schramm, *Biochemistry*, 1997, **36**, 8215-8223.

43. J. Scheuring, P. J. Berti and V. L. Schramm, *Biochemistry*, 1998, **37**, 2748-2758.

44. Y. Tanaka, W. Tao, J. S. Blanchard and E. J. Hehre, *J Biol Chem*, 1994, **269**, 32306-32312.

45. W. Tao, C. Grubmeyer and J. S. Blanchard, *Biochemistry*, 1996, **35**, 14-21.

46. L. A. Reinhardt, D. Svedruzic, C. H. Chang, W. W. Cleland and N. G. J. Richards, *J Am Chem Soc*, 2003, **125**, 1244-1252.

47. C. Bates, Z. Kendrick, N. McDonald and P. C. Kline, *Phytochemistry*, 2006, **67**, 5-12.

48. D. E. Lewis, L. B. Sims, H. Yamataka and J. Mckenna, *J Am Chem Soc*, 1980, **102**, 7411-7419.

49. V. L. Schramm, B. A. Horenstein and P. C. Kline, *J Biol Chem*, 1994, **269**, 18259-18262.

50. V. L. Schramm, *J Biol Chem*, 2007, **282**, 28297-28300.

51. M. J. S. Dewar and G. P. Ford, *J Am Chem Soc*, 1977, **99**, 8343-8344.

52. S. B. Brown, M. J. S. Dewar, G. P. Ford, D. J. Nelson and H. S. Rzepa, *J Am Chem Soc*, 1978, **100**, 7832-7836.

53. J. P. Shea, S. D. Nelson and G. P. Ford, *J Am Chem Soc*, 1983, **105**, 5451-5454.

54. K. Korzekwa, W. Trager, M. Gouterman, D. Spangler and G. H. Loew, *J Am Chem Soc*, 1985, **107**, 4273-4279.

55. I. H. Williams, *J Mol Struct Theochem,* 1983, **94**, 275-284.

56. I. H. Williams, *J Am Chem Soc*, 1984, **106**, 7206-7212.

57. J. Wilkie and I. H. Williams, *J Am Chem Soc*, 1992, **114**, 5423-5425.

58. J. P. Jones and J. L. Urbauer, *J Comput Chem*, 1991, **12**, 1134-1141.

59. J. Andres, V. Moliner and V. S. Safont, *J Chem Soc-Faraday Trans*, 1994, **90**, 1703-1707.

60. J. Andres, V. Moliner, V. S. Safont, L. R. Domingo, M. T. Picher and J. Krechl, *Bioorg Chem*, 1996, **24**, 10-18.

61. E. Gawlita, V. E. Anderson and P. Paneth, *Eur Biophys J*, 1994, **23**, 353-360.

62. J. A. Barnes and I. H. Williams, *Biochem Soc Trans*, 1996, **24**, 263-268.

63. V. Moliner, A. J. Turner and I. H. Williams, *Chem Commun*, 1997, 1271-1272.

64. V. Moliner, J. Andres, M. Oliva, V. S. Safont and O. Tapia, *Theor Chem Acc*, 1999, **101**, 228-233.

65. R. Castillo, J. Andres and V. Moliner, *J Am Chem Soc*, 1999, **121**, 12140-12147.

66. C. Alhambra, J. L. Gao, J. C. Corchado, J. Villa and D. G. Truhlar, *J Am Chem Soc*, 1999, **121**, 2253-2258.

67. J. Rucker and J. P. Klinman, *J Am Chem Soc*, 1999, **121**, 1997-2006.

68. C. Alhambra, J. Corchado, M. L. Sanchez, M. Garcia-Viloca, J. Gao and D. G. Truhlar, *J Phys Chem B*, 2001, **105**, 11326-11340.

69. Q. A. Cui and M. Karplus, *J Phys Chem B*, 2002, **106**, 7927-7947.

70. S. Ferrer, I. Tunon, S. Marti, V. Moliner, M. Garcia-Viloca, A. Gonzalez-Lafont and J. M. Lluch, *J Am Chem Soc*, 2006, **128**, 16851-16863.

71. G. Tresadern, S. Nunez, P. F. Faulder, H. Wang, I. H. Hillier and N. A. Burton, *Faraday Discuss*, 2003, **122**, 223-242.

72. K. E. Ranaghan, L. Masgrau, N. S. Scrutton, M. J. Sutcliffe and A. J. Mulholland, *Chemphyschem*, 2007, **8**, 1816-1835.

73. L. Masgrau, K. E. Ranaghan, N. S. Scrutton, A. J. Mulholland and M. J. Sutcliffe, *J Phys Chem B*, 2007, **111**, 3032-3047.

74. A. Dybala-Defratyka, P. Paneth, R. Banerjee and D. G. Truhlar, *P Natl Acad Sci USA*, 2007, **104**, 10774-10779.

75. A. Dybala-Defratyka, M. Rostkowski and P. Paneth, *Arch Biochem Biophys*, 2008, **474**, 274-282.

76. S. Marti, V. Moliner, I. Tunon and I. H. Williams, *Org Biomol Chem*, 2003, **1**, 483-487.

77. S. E. Worthington, A. E. Roitberg and M. Krauss, *Int J Quantum Chem*, 2003, **94**, 287-292.

78. S. Marti, V. Moliner, M. Tunon and I. H. Williams, *J Phys Chem B*, 2005, **109**, 3707-3710.

79. G. D. Ruggiero, S. J. Guy, S. Marti, V. Moliner and I. H. Williams, *J Phys Org Chem*, 2004, **17**, 592-601.

80. N. Kanaan, J. J. R. Pernia and I. H. Williams, *Chem Commun*, 2008, DOI: 10.1039/b814212b, 6114-6116.

81. J. P. Klinman, *Pure and Applied Chem*, 2003, **75**, 601-608.

82. D. T. Major, A. Heroux, A. M. Orville, M. P. Valley, P. F. Fitzpatrick and J. Gao, *P Natl Acad Sci USA*, 2009, **106**, 20734-20739.

83. Z. D. Nagel and J. P. Klinman, *Nature Chem Biol*, 2009, **5**, 543-550.

84. S. C. L. Kamerlin and A. Warshel, *J. Phys. Org. Chem.*, 2010, **23**, 677-684.

85. A. Warshel, *Abstr Pap Am Chem Soc*, 2012, **244**.

86. J. Zhang and J. P. Klinman, *J Am Chem Soc*, 2011, **133**, 17134-17137.

87. J. P. Klinman, *Nature Chem*, 2010, **2**, 907-909.

88. J. G. Buchanan, G. D. Ruggiero and I. H. Williams, *Org Biomol Chem*, 2008, **6**, 66-72.

89. N. Kanaan, S. Ferrer, S. Marti, M. Garcia-Viloca, A. Kohen and V. Moliner, *J Am Chem Soc*, 2011, **133**, 6692-6702.

90. G. D. Ruggiero, I. H. Williams, M. Roca, V. Moliner and I. Tunon, *J Am Chem Soc*, 2004, **126**, 8634-8635.

91. J. Y. Zhang and J. P. Klinman, *J Am Chem Soc*, 2011, **133**, 17134-17137.

92. J. Zhang, H. J. Kulik, T. J. Martinez and J. P. Klinman, *P Natl Acad Sci USA*, 2015, **112**, 7954-7959.

93. V. Moliner and I. H. Williams, *J Am Chem Soc*, 2000, **122**, 10895-10902.

94. R. A. Kwiecien, J.-Y. Le Questel, J. Lebreton, M. Delaforge, F. Andre, E. Pihan, A. Roussel, A. Fournial, P. Paneth and R. J. Robins, *J Phys Chem B*, 2012, **116**, 7827-7840.

95. A. Siwek, R. Omi, K. Hirotsu, K. Jitsumori, N. Esaki, T. Kurihara and P. Paneth, *Arch Biochem Biophys*, 2013, **540**, 26-32.

96. A. Krzeminska, P. Paneth, V. Moliner and K. Swiderek, *J Phys Chem B*, 2015, **119**, 917-927.

97. K. Swiderek, A. Kohen and V. Moliner, *Phys Chem Chem Phys*, 2015, **17**, 30793-30804.

98. F. Jensen, *Introduction to computational chemistry*, John Wiley & Sons, 2013.

99. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *J Comput Chem*, 1983, **4**, 187-217.

100. K. K. Irikura, R. D. Johnson, III, R. N. Kacker and R. Kessel, *J Chem Phys*, 2009, **130**.

101. M. Born and R. Oppenheimer, *Annalen der Physik*, 1927, **389**, 457-484.

102. W. Kohn and L. J. Sham, *Phys Rev*, 1965, **140**, 1133-1138.

103. W. Kohn, *Int J Quantum Chem*, 1995, **56**, 229-232.

104. J. P. Perdew, *Physica B*, 1991, **172**, 1-6.

105. C. Møller and M. S. Plesset, *Physical Review*, 1934, **46**, 618-622.

106. R. J. Bartlett, *J Phys Chem*, 1989, **93**, 1697-1708.

107. W. Koch, M. C. Holthausen and E. J. Baerends, *A Chemist's Guide to Density Functional Theory*, FVA-Frankfurter Verlagsanstalt GmbH, 2001.

108. C. W. Bauschlicher and H. Partridge, *J Chem Phys*, 1995, **103**, 1788-1791.

109. J. Simons, *Introduction to Theoretical Chemistry*, Cambridge University Press, Cambridge, 2003.

110. Y. Alexeev, M. P. Mazanetz, O. Ichihara and D. G. Fedorov, *Curr Top Med Chem*, 2012, **12**, 2013-2033.

111. Y. Tantirungrotechai, K. Phanasant, S. Roddecha, P. Surawatanawong, V. Sutthikhum and J. Limtrakul, *J Mol Struct Theochem*, 2006, **760**, 189-192.

112. J. P. Perdew, A. Ruzsinszky, J. Tao, V. N. Staroverov, G. E. Scuseria and G. I. Csonka, *J Chem Phys*, 2005, **123**, 062201.

113. S. H. Vosko, L. Wilk and M. Nusair, *Canad J Phys*, 1980, **58**, 1200-1211.

114. S. Grimme, *J Comput Chem*, 2006, **27**, 1787-1799.

115. Y. Zhao and D. G. Truhlar, *J Phys Chem A*, 2004, **108**, 6908-6918.

116. C. Adamo and V. Barone, *J Chem Phys*, 1999, **110**, 6158-6170.

117. R. Car, *Nature Chem*, 2016, **8**, 820-821.

118. J. B. Foresman and A. E. Frisch, *Exploring chemistry with electronic structure methods*, Gaussian, Inc., 1996.

119. J. N. Israelachvili, in *Intermolecular and Surface Forces (Third Edition)*, Academic Press, San Diego, 2011, pp. 107-132.

120. K. Szalewicz, *Wires Comput Mol Sci*, 2012, **2**, 254-272.

121. S. Grimme, A. Hansen, J. G. Brandenburg and C. Bannwarth, *Chem Rev*, 2016, **116**, 5105-5154.

122. S. Rosel, H. Quanz, C. Logemann, J. Becker, E. Mossou, L. Canadillas-Delgado, E. Caldeweyher, S. Grimme and P. R. Schreiner, *J Am Chem Soc*, 2017, **139**, 7428-7431.

123. G. A. Petersson and M. A. Allaham, *J Chem Phys*, 1991, **94**, 6081-6090.

124. J. C. Slater, *The Self-consistent Field for Molecules and Solids*, McGraw-Hill, 1974.

125. J. S. Binkley, J. A. Pople and W. J. Hehre, *J Am Chem Soc*, 1980, **102**, 939-947.

126. T. Clark, J. Chandrasekhar, G. W. Spitznagel and P. V. Schleyer, *J Comput Chem*, 1983, **4**, 294-301.

127. I. H. Williams, P. H. Sugden and H. E. Morgan, *Am J Physiol*, 1981, **240**, E677-E681.

128. J. Tomasi, B. Mennucci and E. Cances, *J Mol Struct Theochem*, 1999, **464**, 211-226.

129. A. Warshel, M. Levitt and S. Lifson, *J Mol Spectrosc*, 1970, **33**, 84.

130. M. Levitt and A. Warshel, *Nature*, 1975, **253**, 694-698.

131. A. Warshel and M. Levitt, *J Mol Biol*, 1976, **103**, 227-249.

132. A. Warshel and M. Levitt, *J Mol Biol*, 1976, **106**, 421-437.

133. M. Levitt and A. Warshel, *J Am Chem Soc*, 1978, **100**, 2607-2613.

134. M. J. Field, P. A. Bash and M. Karplus, *J Comput Chem*, 1990, **11**, 700-733.

135. H. M. Senn and W. Thiel, *Atomistic Approaches in Modern Biology: From Quantum Chemistry to Molecular Simulations*, 2007, **268**, 173-290.

136. P. Sherwood, B. R. Brooks and M. S. P. Sansom, *Curr Opin Struc Biol*, 2008, **18**, 630-640.

137. D. Riccardi, S. Yang and Q. Cui, *Biochi Biophys Acta, Proteins Proteomics*, 2010, **1804**, 342-351.

138. W. Yang, Q. Cui, D. H. Min and H. Z. Li, *Ann Rep Comp Chem*, 2010, **6**, 51-62.

139. C. L. Avila, N. J. D. Drechsel, R. Alcantara and J. Villa-Freixa, *Curr Protein Pept Sc*, 2011, **12**, 221-234.

140. R. C. Braga and C. H. Andrade, *Mini-Rev Med Chem*, 2012, **12**, 573-582.

141. O. Acevedo, *J Phys Chem A*, 2014, **118**, 11653-11666.

142. O. Acevedo and W. L. Jorgensen, *Wires Comput Mol Sci*, 2014, **4**, 422-435.

143. M. Y. Liu, Y. J. Wang, Y. K. Chen, M. J. Field and J. L. Gao, *Isr J Chem,* 2014, **54**, 1250-1263.

144. F. Duarte, B. A. Amrein, D. Blaha-Nelson and S. C. L. Kamerlin, *Biochim Biophys Acta, Gen Subjects*, 2015, **1850**, 954-965.

145. T. Borowski, M. Quesne and M. Szaleniec, *Combined Quantum Mechanical and Molecular Mechanical Modelling of Biomolecular Interactions*, 2015, **100**, 187-224.

146. X. Y. Lu, D. Fang, S. Ito, Y. Okamoto, V. Ovchinnikov and Q. Cui, *Mol Sim*, 2016, **42**, 1056-1078.

147. M. Zheng and M. P. Waller, *Wiley Interdiscip Rev Comput Mol Sci*, 2016, **6**, 369-385.

148. S. F. Sousa, A. J. M. Ribeiro, R. P. P. Neves, N. F. Bras, N. M. F. S. A. Cerqueira, P. A. Fernandes and M. J. Ramos, *Wiley Interdiscip Rev Comput Mol Sci,* 2017, **7,** doi: 10.1002/wcms.1281.

149. K. E. Ranaghan and A. J. Mulholland, *Int Rev Phys Chem*, 2010, **29**, 65-133.

150. M. W. van der Kamp and A. J. Mulholland, *Biochemistry*, 2013, **52**, 2708-2728.

151. J. Gao, S. Ma, D. T. Major, K. Nam, J. Pu and D. G. Truhlar, *Chem Rev*, 2006, **106**, 3188-3209.

152. H. M. Senn and W. Thiel, *Angew Chem Int Ed*, 2009, **48**, 1198-1229.

153. G. Groenhof, *Biomolecular simulations: methods and protocols*, 2013, 43-66.

154. F. Maseras and K. Morokuma, *Chem Phys Lett*, 1992, **195**, 500-504.

155. F. Maseras, X. K. Li, N. Koga and K. Morokuma, *J Am Chem Soc*, 1993, **115**, 10974-10980.

156. F. Maseras, N. Koga and K. Morokuma, *J Am Chem Soc*, 1993, **115**, 8313-8320.

157. F. Maseras, N. Koga and K. Morokuma, *Organometallics*, 1994, **13**, 4008-4016.

158. F. Maseras and K. Morokuma, *J Comput Chem*, 1995, **16**, 1170-1179.

159. M. Svensson, S. Humbel, R. D. J. Froese, T. Matsubara, S. Sieber and K. Morokuma, *J Phys Chem*, 1996, **100**, 19357-19363.

160. K. Doitomi and H. Hirao, *Tetrahedron Lett*, 2017, **58**, 2309-2317.

161. C. L. Firme, N. K. V. Monteiro and S. R. B. Silva, *Comput Theor Chem*, 2017, **1111**, 40-49.

162. P. Kostetskyy and G. Mpourmpakis, *Ind Eng Chem Res*, 2017, **56**, 7062-7069.

163. R. P. P. Neves, P. A. Fernandes and M. J. Ramos, *P Natl Acad Sci USA*, 2017, **114**, E4724-E4733.

164. J. M. Wang, Y. N. Zuo, C. W. Hu and Z. S. Su, *Catal Sci Technol*, 2017, **7**, 2183-2193.

165. L. W. Chung, W. M. C. Sameera, R. Ramozzi, A. J. Page, M. Hatanaka, G. P. Petrova, T. V. Harris, X. Li, Z. Ke, F. Liu, H.-B. Li, L. Ding and K. Morokuma, *Chem Rev*, 2015, **115**, 5678-5796.

166. D. M. Philipp and R. A. Friesner, *J Comput Chem*, 1999, **20**, 1468-1494.

167. R. B. Murphy, D. M. Philipp and R. A. Friesner, *Chem Phys Lett*, 2000, **321**, 113-120.

168. M. Wirstam, S. J. Lippard and R. A. Friesner, *J Am Chem Soc*, 2003, **125**, 3980-3987.

169. B. F. Gherman, S. J. Lippard and R. A. Friesner, *J Am Chem Soc*, 2005, **127**, 1025-1037.

170. D. Rinaldo, D. M. Philipp, S. J. Lippard and R. A. Friesner, *J Am Chem Soc*, 2007, **129**, 3135-3147.

171. T. F. Hughes and R. A. Friesner, *J Chem Theory Comput*, 2012, **8**, 442-459.

172. T. F. Hughes, J. N. Harvey and R. A. Friesner, *Phys Chem Chem Phys*, 2012, **14**, 7724-7738.

173. C. L. McMullin, J. Jover, J. N. Harvey and N. Fey, *Dalton T*, 2010, **39**, 10833-10836.

174. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, Gaussian 09 (Revision D.01); Gaussian Inc.: Wallingford, CT, **2009**.

175. I. H. Williams, J. Mckenna and L. B. Sims, *J Mol Struct*, 1979, **55**, 147-150.

176. I. H. Williams, *J Mol Struct Theochem*, 1983, **11**, 275-284.

177. D. A. Stauffer, R. E. Barrans, and D. A. Dougherty, *Angew Chem Int Edit*, **29**(8), 915-918.

178. J. Bigeleisen, M. W. Lee and F. Mandel, *Ann Rev Phys Chem*, 1973, **24**, 407-440.

179. J. Bron, C. F. Chang and M. Wolfsberg, *Zeitschrift Fur Naturforschung Section a: Journal of Physical Sciences*, 1973, **A 28**, 129-136.

180. S. Padrao, S. M. Fiuza, A. M. Amado, A. M. A. da Costa and L. A. E. B. de Carvalho, *J. Phys. Org. Chem.*, 2011, **24**, 110-121.

181. A. M. Amado, S. M. Fiuza, L. A. E. B. de Carvalho and P. J. A. Ribeiro-Claro, *B Chem Soc Jpn*, 2012, **85**, 962-975.

182. A. M. Amado, S. M. Fiuza, L. A. E. B. de Carvalho and P. J. A. Ribeiro-Claro, *J Chem*, 2013, DOI: 10.1155/2013/682514.

183. M. Born, *Der aufbau der materie*, J. Springer, Berlin,, 1920.

184. M. Born, *Die relativitätstheorie Einsteins und ihre physikalischen grundlagen gemeinverständlich dargestellt von Max Born*, J. Springer, Berlin,, 1920.

185. A. V. Marenich, C. J. Cramer and D. G. Truhlar, *J Phys Chem B*, 2009, **113**, 6378-6396.

186. I. H. Williams, *J Chem Soc Chem Comm*, 1985, DOI 10.1039/c39850000510, 510-511.

187. M. Wolfsberg, *J Chem Phys*, 1969, **50**, 1484.

188. M. Wolfsberg, W. A. Van Hook, P. Paneth, L. P. N. Rebelo, M. Wolfsberg, W. A. VanHook and P. Paneth, *Enzymes; Aqueous Solvent IE'S*, 2010.

189. D. J. O'Leary, D. D. Hickstein, B. K. V. Hansen and P. E. Hansen, *J. Org. Chem.*, 2010, **75**, 1331-1342.

190. R. L. Jacobsen, R. D. Johnson, K. K. Irikura and R. N. Kacker, *J Chem Theory Comput*, 2013, **9**, 951-954.

191. E. Roueff, M. Gerin, D. C. Lis, A. Wootten, N. Marcelino, J. Cernicharo and B. Tercero, *J Phys Chem A*, 2013, **117**, 9959-9967.

192. A. D. Isaacson and D. G. Truhlar, *J Chem Phys*, 1981, **75**, 4090-4094.

193. A. D. Isaacson, D. G. Truhlar, K. Scanlon and J. Overend, *J Chem Phys*, 1981, **75**, 3017-3024.

194. B. K. C. de Miranda, C. Alcaraz, M. Elhanine, B. Noller, P. Hemberger, I. Fischer, G. A. Garcia, H. Soldi-Lose, B. Gans, L. A. V. Mendes, S. Boye-Peronne, S. Douin, J. Zabka and P. Botschwina, *J Phys Chem A*, 2010, **114**, 4818-4830.

195. V. Barone, *J Chem Phys*, 2004, **120**, 3059-3065.

196. V. Barone, P. Carbonniere, and C. Pouchan, *J Chem Phys*, 2005, **122**, 224308.

197. M. W. D. Hanson-Heine, M. W. George and N. A. Besley, *J Chem Phys*, 2012, **136**.

198. M. W. D. Hanson-Heine, M. W. George and N. A. Besley, *J Phys Chem A*, 2012, **116**, 4417-4425.

199. A. S. Menon, G. P. F. Wood, D. Moran and L. Radom, *J Phys Chem A*, 2007, **111**, 13638-13644.

200. J. P. Merrick, D. Moran and L. Radom, *J Phys Chem A*, 2007, **111**, 11683-11700.

201. M. Roca, S. Marti, J. Andres, V. Moliner, I. Tunon, J. Bertran and I. H. Williams, *J Am Chem Soc*, 2003, **125**, 7726-7737.

202. J. Bigeleisen, *J Chem Phys*, 1955, **23**, 2264-2267.

203. C. Cappelli, S. Monti, G. Scalmani and V. Barone, *J Chem Theory Comput*, 2010, **6**, 1660-1669.

204. S. Horowitz, L. M. A. Dirk, J. D. Yesselman, J. S. Nimtz, U. Adhikari, R. A. Mehl, S. Scheiner, R. L. Houtz, H. M. Al-Hashimi and R. C. Trievel, *J Am Chem Soc*, 2013, **135**, 15536-15548.

205. J. Y. Zhang and J. P. Klinman, *J Am Chem Soc*, 2016, **138**, 9158-9165.

206. J. P. Wang, Y. Pan, P. Li and G. J. Zhao, *J Alloy Compd*, 2016, **686**, 656-661.

207. T. Schmidt, T. Schwede and M. Meuwly, *J Phys Chem B*, 2014, **118**, 5882-5890.

208. M. F. Hegazi, R. T. Borchardt and R. L. Schowen, *J Am Chem Soc*, 1979, **101**, 4359-4365.

209. J. Y. Zhang, H. J. Kulik, T. J. Martinez and J. P. Klinman, *Proc Natl Acad Sci USA*, 2015, **112**, 7954-7959.

210. Y. Fazli, H. Alijani and K. Khezri, *Adv Polym Tech*, 2016, **35**, 260-268.

211. L. Zhang, X. Gao, Z. X. Zhang, M. B. Zhang, Y. Q. Cheng and J. X. Su, *Sci Rep-Uk*, 2016, **6**.

212. D. Z. Wu, L. Zhao, V. K. Vakharia, W. Salim and W. S. W. Ho, *J Membrane Sci*, 2016, **510**, 58-71.

213. Z. Z. Feng, J. J. Yu, J. M. Kong and T. H. Wang, *Chem Eng J*, 2016, **294**, 236-245.

214. L. Yang, M. G. Gao, B. Dai, X. H. Guo, Z. Y. Liu and B. H. Peng, *Appl Surf Sci*, 2016, **386**, 337-344.

215. C. Pirola, F. Galli, M. Corbetta and F. Manenti, *Clean Technol Envir*, 2015, **17**, 1139-1147.

216. G. B. Yan, A. J. Borah, L. G. Wang and M. H. Yang, *Adv Synth Catal*, 2015, **357**, 1333-1350.

217. T. M. Gogsig, J. Kleimark, S. O. N. Lill, S. Korsager, A. T. Lindhardt, P. O. Norrby and T. Skrydstrup, *J Am Chem Soc*, 2012, **134**, 443-452.

218. G. A. Craze, A. J. Kirby and R. Osborne, *J Chem Soc Perk T 2*, 1978, 357-369.

219. B. L. Knier and W. P. Jencks, *J Am Chem Soc*, 1980, **102**, 6789-6798.

220. G. D. Ruggiero and I. H. Williams, *J Chem Soc Perk T 2*, 2001, 448-458.

221. S. Wolfe, C. K. Kim, K. Y. Yang, N. Weinberg and Z. Shi, *Can J Chem*, 1998, **76**, 359-370.

222. M. C. Thielges, D. A. Case and F. E. Romesberg, *J Am Chem Soc*, 2008, **130**, 6597-6603.

223. V. L. Schramm, *ACS Chem Biol*, 2013, **8**, 71-81.

224. R. G. Silva, A. S. Murkin and V. L. Schramm, *Proc Natl Acad Sci U S A*, 2011, **108**, 18661-18665.

225. I. H. Williams, *J Chem Theory Comput*, 2012, **8**, 542-553.

226. T. D. Poulsen, M. Garcia-Viloca, J. L. Gao and D. G. Truhlar, *J Phys Chem B*, 2003, **107**, 9567-9578.

227. L. Masgrau and D. G. Truhlar, *Acc Chem Res*, 2015, **48**, 431-438.

228. M. J. Stern and M. Wolfsberg, *J Pharm Sci*, 1965, **54**, 849-858.

229. G. D. Ruggiero and I. H. Williams, *J Chem Soc Perk T 2*, 2002, 591-597.

230. C. G. Swain, E. C. Stivers, J. F. Reuwer and L. J. Schaad, *J Am Chem Soc*, 1958, **80**, 5885-5893.

231. A. J. Turner, V. Moliner and I H. Williams, *Phys. Chem. Chem. Phys.*, 1999, **1**, 1323-1331.

232. J. K. Hwang, Z. T. Chu, A. Yadav and A. Warshel, *J Phys Chem-Us*, 1991, **95**, 8445-8448.

233. J. J. Ruiz-Pernia, M. Garcia-Viloca, S. Bhattacharyya, J. L. Gao, D. G. Truhlar and I. Tunon, *J Am Chem Soc*, 2009, **131**, 2687-2698.

234. J. L. Bao, R. Meana-Paneda and D. G. Truhlar, *Chem Sci*, 2015, **6**, 5866-5881.

235. J. J. R. Pernia and I. H. Williams, *Chem Eur J*, 2012, **18**, 9405-9414.

236. D. G. Truhlar, *Arch Biochem Biophys,* 2015, **582**, 10-17.

237. A. P. Nyczepir, A. K. Nagel and G. Schnabel, *Hortscience*, 2009, **44**, 1932-1935.

238. K. E. Ranaghan and A. J. Mulholland, *Interdiscip Sci,* 2010, **2**, 78-97.

239. K. Swiderek, J. J. Ruiz-Pernia, V. Moliner and I. Tunon, *Curr Opin Chem Biol*, 2014, **21**, 11-18.

240. K. Zinovjev and I. Tunon, *J Chem Phys*, 2015, **143**, 134111.

241. I. R. Greig and I. H. Williams, *Chem Commun*, 2007, 3747-3749.

242. J. L. Gao, K. Y. Wong and D. T. Major, *J Comput Chem*, 2008, **29**, 514-522.

243. A. G. Algarra, W. B. Cross, D. L. Davies, Q. Khamker, S. A. Macgregor, C. L. McMullin and K. Singh, *J. Org. Chem.*, 2014, **79**, 1954-1970.

244. F. London, *Trans Faraday Soc*, 1937, **33**, 8b-26.

245. A. Borba, M. Albrecht, A. Gomez-Zavaglia, M. A. Suhm and R. Fausto, *J Phys Chem A*, 2010, **114**, 151-161.

246. K. S. Pitzer and E. Catalano, *J Am Chem Soc*, 1956, **78**, 4844-4846.

247. S. O. N. Lill, P. Ryberg, T. Rein, E. Bennstrom and P. O. Norrby, *Chem Eur J,* 2012, **18**, 1640-1649.

248. E. R. Davidson and D. Feller, *Chem Rev*, 1986, **86**, 681-696.

249. T. H. Dunning, *J Chem Phys*, 1989, **90**, 1007-1023.

250. R. A. Kendall, T. H. Dunning and R. J. Harrison, *J Chem Phys*, 1992, **96**, 6796-6806.

251. S. Kristyan and P. Pulay, *Chem Phys Lett*, 1994, **229**, 175-180.

252. Y. J. Zhu, M. Bauer, J. Ploog and L. Ackermann, *Chem Eur J*, 2014, **20**, 13099-13102.

253. H. Kruse, L. Goerigk and S. Grimme, *J. Org. Chem.*, 2012, **77**, 10824-10834.

254. R. Lonsdale, J. N. Harvey and A. J. Mulholland, *Chem Soc Rev*, 2012, **41**, 3025-3038.

255. R. Lonsdale, J. N. Harvey and A. J. Mulholland, *J Phys Chem Lett,* 2010, **1**, 3232-3237.

256. M. W. van der Kamp, J. Zurek, F. R. Manby, J. N. Harvey and A. J. Mulholland, *J Phys Chem B*, 2010, **114**, 11303-11314.

257. S. P. de Visser, M. G. Quesne, B. Martin, P. Comba and U. Ryde, *Chem Commun*, 2014, **50**, 262-282.

258. M. Swart, M. Sola and F. M. Bickelhaupt, *J Comput Chem*, 2011, **32**, 1117-1127.

259. M. Swart and P. T. Van Duijnen, *Int J Quantum Chem*, 2011, **111**, 1763-1772.

260. M. Swart, M. Sola and F. M. Bickelhaupt, *J Chem Phys*, 2009, **131**.

261. R. P. Matthews, T. Welton and P. A. Hunt, *Phys Chem Chem Phys*, 2014, **16**, 3238-3253.

262. M. Roca, V. Moliner, J. J. Ruiz-Pernia, E. Silla and I. Tunon, *J Phys Chem A*, 2006, **110**, 503-509.

263. B. Civalleri, C. M. Zicovich-Wilson, L. Valenzano and P. Ugliengo, *Crystengcomm*, 2008, **10**, 405-410.

264. L. Perrin, K. J. T. Carr, D. McKay, C. L. McMullin, S. A. Macgregor and O. Eisenstein, *Struct Bond*, 2016, **167**, 1-37.

265. D. L. Davies, C. E. Ellul, S. A. Macgregor, C. L. McMullin and K. Singh, *J Am Chem Soc*, 2015, **137**, 9659-9669.

266. A. G. Algarra, D. L. Davies, Q. Khamker, S. A. Macgregor, C. L. McMullin, K. Singh and B. Villa-Marcos, *Chem Eur J,* 2015, **21**, 3087-3096.

267. D. L. Davies, C. E. Ellul, S. A. Macgregor, C. L. McMullin and K. Singh, *J Am Chem Soc*, 2015, **137**, 9659-9669.

268. A. G. Algarra, D. L. Davies, Q. Khamker, S. A. Macgregor, C. L. McMullin, K. Singh and B. Villa-Marcos, *Chemistry*, 2015, **21**, 3087-3096.

269. R. H. Mei, H. Wang, S. Warratz, S. A. Macgregor and L. Ackermann, *Chem Eur J* 2016, **22**, 6759-6763.

270. K. Raghuvanshi, D. Zell, K. Rauch and L. Ackermann, *ACS Catal*, 2016, **6**, 3172-3175.

271. Z. X. Ruan, S. Lackner and L. Ackermann, *Angew Chem Int Edit*, 2016, **55**, 3153-3157.

272. M. Moselage, J. Li and L. Ackermann, *ACS Catal*, 2016, **6**, 498-525.

273. R. H. Mei, J. Loup and L. Ackermann, *ACS Catal*, 2016, **6**, 793-797.

274. G. Cera, T. Haven and L. Ackermann, *Angew Chem Int Edit*, 2016, **55**, 1484-1488.

275. D. Zell, S. Warratz, D. Gelman, S. J. Garden and L. Ackermann, *Chem Eur J,* 2016, **22**, 1248-1252.

276. A. Bechtoldt, C. Tirler, K. Raghuvanshi, S. Warratz, C. Kornhaass and L. Ackermann, *Angew Chem Int Edit*, 2016, **55**, 264-267.

277. W. B. Ma and L. Ackermann, *ACS Catal*, 2015, **5**, 2822-2825.

278. S. Warratz, C. Kornhaass, A. Cajaraville, B. Niepotter, D. Stalke and L. Ackermann, *Angew Chem Int Edit*, 2015, **54**, 5513-5517.

279. J. Li and L. Ackermann, *Chem Eur J*, 2015, **21**, 5718-5722.

280. W. P. Liu, D. Zell, M. John and L. Ackermann, *Angew Chem Int Edit*, 2015, **54**, 4092-4096.

281. J. Li and L. Ackermann, *Angew Chem Int Edit*, 2015, **54**, 3635-3638.

282. S. Kozuch and J. M. L. Martin, *Chemphyschem*, 2011, **12**, 1413-1418.

283. A. Uhe, S. Kozuch and S. Shaik, *J Comput Chem*, 2011, **32**, 978-985.

284. A. D. Becke, *Phys Rev A*, 1988, **38**, 3098-3100.

285. M. Levy and J. P. Perdew, *Phys Rev A*, 1985, **32**, 2010.

286. M. Dolg, U. Wedig, H. Stoll and H. Preuss, *J Chem Phys*, 1987, **86**, 866-872.

287. V. A. Rassolov, M. A. Ratner, J. A. Pople, P. C. Redfern and L. A. Curtiss, *J Comput Chem*, 2001, **22**, 976-984.

288. V. A. Rassolov, J. A. Pople, M. A. Ratner and T. L. Windus, *J Chem Phys*, 1998, **109**, 1223-1229.

289. S. Grimme, S. Ehrlich and L. Goerigk, *J Comput Chem*, 2011, **32**, 1456-1465.

290. A. G. Algarra, W. B. Cross, D. L. Davies, Q. Khamker, S. A. Macgregor, C. L. McMullin and K. Singh, *J Org Chem*, 2014, **79**, 1954-1970.

291. G. N. Patargias, S. A. Harris and J. H. Harding, *J Chem Phys*, 2010, **132**(23), 235103.

292. F. Egidi, V. Barone, J. Bloino and C. Cappelli, *J Chem Theory Comput*, 2012, **8**, 585-597.

293. I. H. Williams and P. B. Wilson, *SoftwareX*, 2017, **6**, 1-6.

294. H. J. Kulik, J. Zhang, J. P. Klinman and T. J. Martínez, *J Phys Chem B*, 2016, **120**, 11381-11394.

295. G. Jindal and A. Warshel, *J Phys Chem B*, 2016, **120**, 9913-9921.